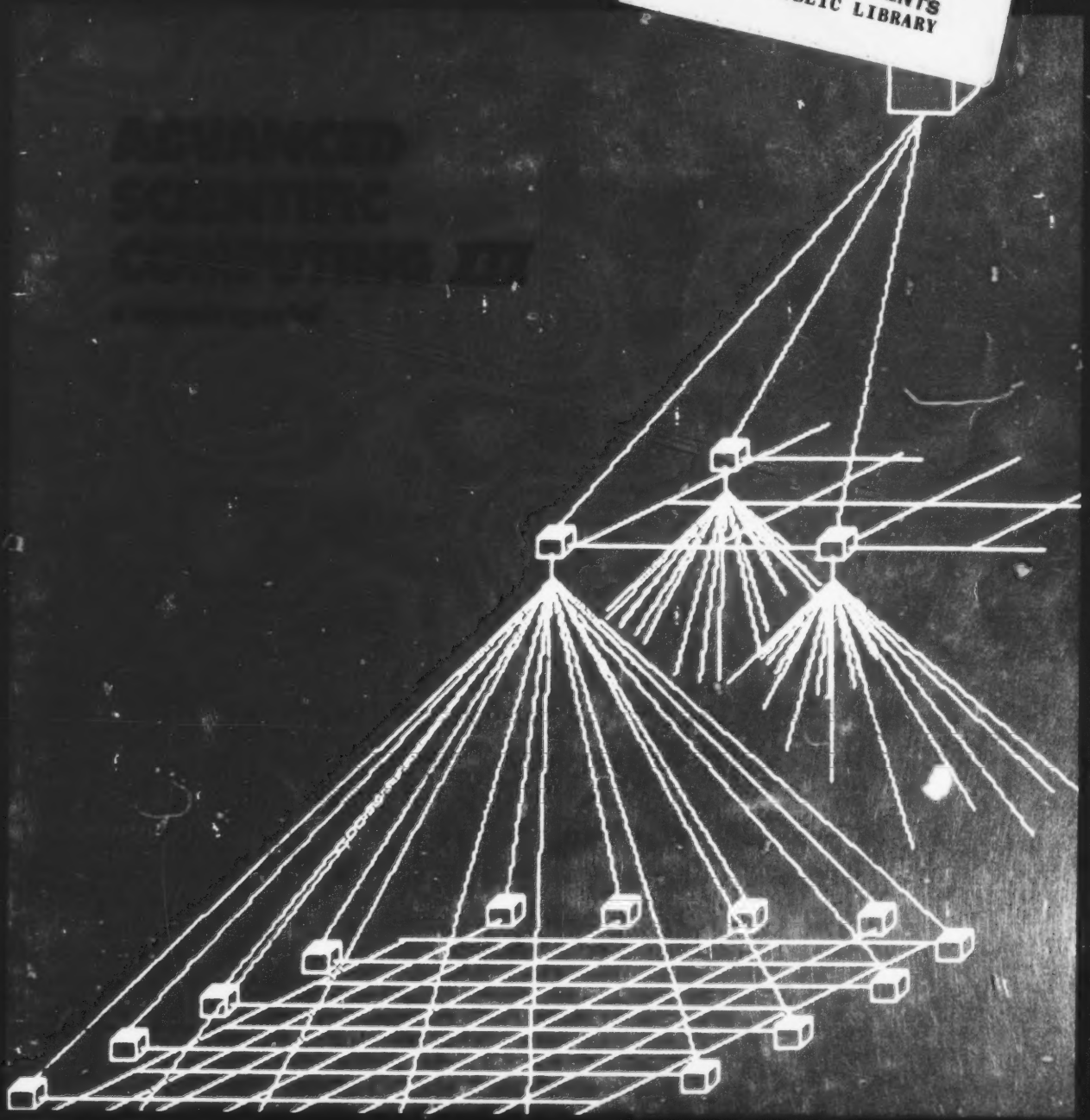
.17 /1

NATIONAL SCIENCE FOUNDATION

Many-body problems
Neuroscience
Cognitive sciences
Imaging
Economics

VOLUME 17 NUMBER 1 · SPRING 1986

LD [redacted] C
U.S. DOCUMENTS
NEWARK PUBLIC LIBRARY



NS1.29:

In this Mosaic

With this issue, *Mosaic* completes the most ambitious project of its 16 years: the consideration of the issues and the prospects that make up the field of advanced scientific computing. What began as an article became a special issue of the magazine; those bounds, too, were soon burst, and a three-issue special, Volume 16 Numbers 3 and 4 and Volume 17 Number 1, was born. The first issue dealt with the subject from the perspective of computer science and engineering, the second and third with the disciplinary and interdisciplinary problems driving the development. (For those who missed them, the first two can be ordered as back issues with the enclosed subscription card.)

To have dealt with advanced scientific computing—even in three magazines and some 70,000 words—is not to have left it behind. If we showed anything in the effort, it is that computing in many ways is both the shape and the shaper of research to come. No field is likely to be exempt from its effects, not only in defining the way problems will be attacked and solved, but even in describing the problems themselves that are amenable to solution.

We think the product was well worth the effort. We are massively indebted not only to the authors whose work the magazine really is, but also to the investigators, without whose cooperation we could never achieve what we do.

Warren Kornberg

MOSAIC

The National Science Foundation is charged to help stretch and redefine research frontiers. It must also communicate with the scientists, engineers, administrators, educators, and students at work along those frontiers. Part of that job falls to the magazine *Mosaic*, a uniquely readable and reliable account of current work and thought in the vast array of research areas with which NSF is concerned. *Mosaic* has shown itself to be valuable to those who perform, teach, study, or otherwise care about scientific and engineering research. It is also generally available. To receive a sample copy, write: Distribution, *Mosaic* (533), NSF, Washington, D.C. 20550. To subscribe, turn this card over.

NEWARK PUBLIC LIBRARY

JUN 2 1986

U. S. GOVT. PUBS.
DEPOSITORY 378

NATIONAL SCIENCE FOUNDATION

Roland W. Schmitt, *Chairman*
National Science Board

Erich Bloch, *Director*
National Science Foundation

Raymond E. Bye, Jr., *Director*
Office of Legislative
and Public Affairs

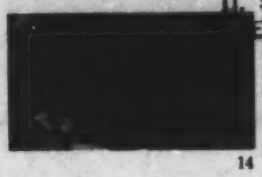
Warren Kornberg, *Editor*

Mosaic is published as a source of information for scientific and educational communities served by the National Science Foundation, and the general public. The Director of the National Science Foundation has determined that the publication of this periodical is necessary in the transaction of the public business required by law of this agency. Use of funds for printing this periodical has been approved by the Director of the Office of Management and Budget through March 31, 1991. Subscriptions are available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Subscription is $15.00 per year in the United States and possessions and $18.75 in all other countries. Single copies are $2.75 domestic and $3.44 foreign. Send check or money order to the U.S. Government Printing Office. Communications other than those concerning paid subscriptions should be addressed to: Editor, *Mosaic*, National Science Foundation, Washington, D.C. 20550. Telephone: 202-357-9838.

Illustration credits: Cover, 6—Lawrence Berkeley Laboratory; 2—P. Hut; 4—Drexel U.; 5, 6—UC, Santa Barbara, Scalapino; 7, 8—Cohen, Louie; 9—Vanderbilt U.; 10—Harvard-Smithsonian Center for Astrophysics, Mary Creswell; 11—C. F. Fischer; 12, 13—S. Koonin; 14—K. Johnson; 15—Columbia U.; 16—T. Poggio, MIT; 17, 42—U. of Pennsylvania, G. Gerstein; 19—Johns Hopkins U.; 20, 21—D. Marr: Vision, W. H. Freeman and Co. Publishers; 22, 23, 36—I. Bohachevsky, UCLA; 24, 45—Carnegie-Mellon U.; 25—UC, San Diego; 26—E. Tzanakou, Pat Bacon; 27, 30—Stanford U.; 31, 43—Edward W. Souza, Rick Friedman; 32—Caltech; 33—R. Clayton; 34—G. Herman, U. of Pennsylvania; 35—M. Phelps; 36—J. A. Lake; 37—U. Aebi, J. Brough Schamp; 40—A. Steven, NIH; 41—NCAR; 44—R. Fogel; 45—U. of Chicago; 46, 47—V. Smith, Indiana U.

Cover: Conceptual organization of processors and the flow of information in *Midas*.

2



14

32

Many-body Problems *by Mort La Brecque* 2
Once the number of bodies grows beyond two, the problems of computing their effects on each other begin to grow out of control.
One physicist's simulation 9

Taunting the Computer *by Norman Metzger* 14
The chimerical similarities between neurobiology and computing continue to dissolve on approach. But the fields do support each other in many ways.
Computer vision 18
A Marr sampler 20

Cognition and Computation *by Arthur Fisher* 22
The massively parallel architectures being devised for scientific computing promise cognitive scientists for the first time a set of tools worthy of the problems demanding attention.
Speech recognition 28

A Picture is Worth . . . *by William Check* 32
Sometimes images are the only way a problem or solution can be represented. Getting images into and out of computing systems is a constantly advancing frontier.
Architecture and algorithms 38

The Problem of Macrovariables *by Edward Edelson* 42
Many economists, like their colleagues in the ocean, earth, and atmospheric sciences, deal with systems so massively fluid as to defy—so far—efforts to model them precisely.

Letters, etc. **Inside back cover**

Federal Funds for Mathematics and Computer Sciences Growing Sharply **Back cover**

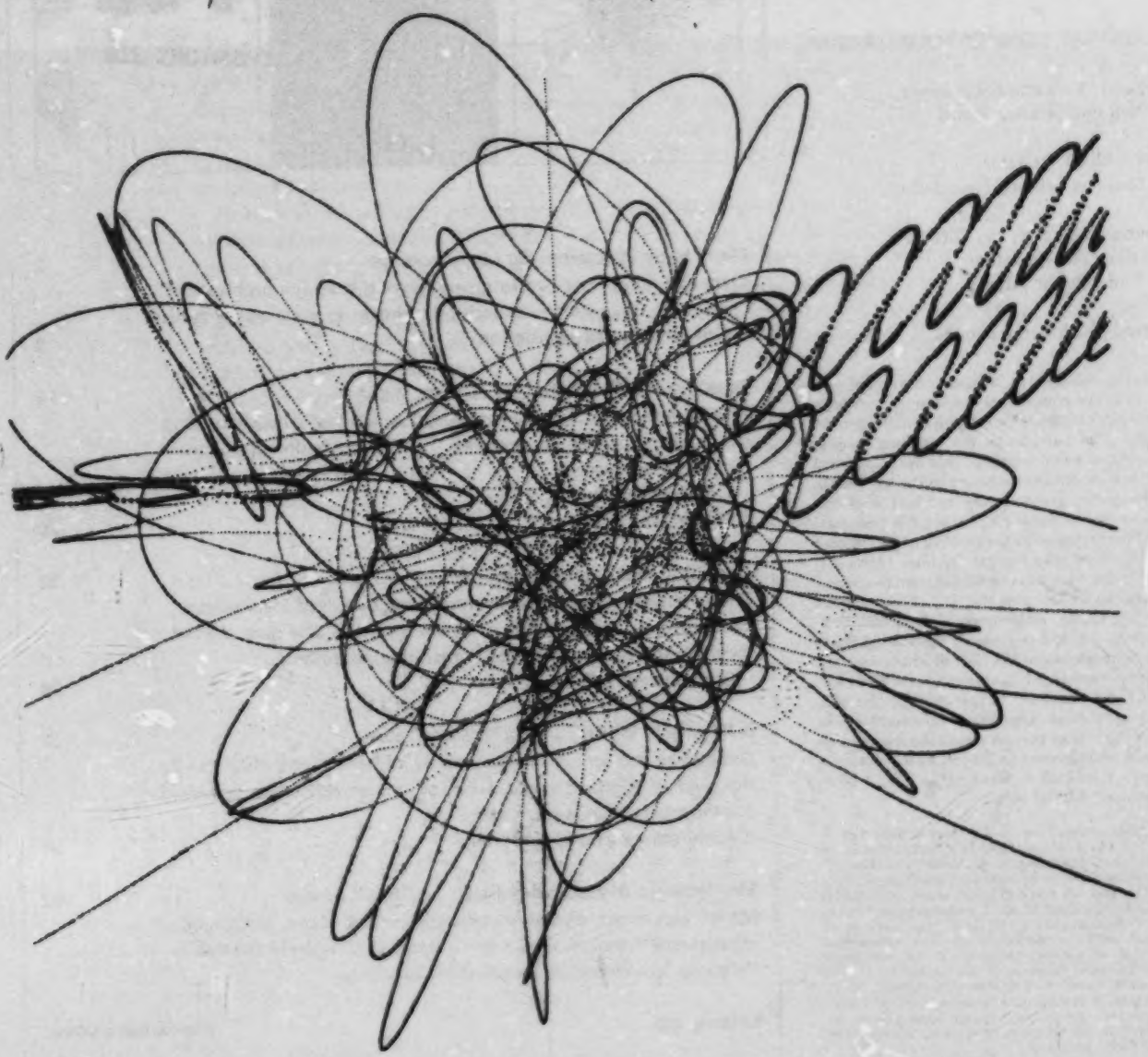
READEX

best copy available

MANY-BODY PROBLEMS

by Mort La Brecque

Once the number of bodies interacting with each other grows beyond two, the problems of computing their effects on each other begin to grow out of control.



Scientific discoveries in the twentieth century have vindicated Democritus, the ancient Greek philosopher whose atomistic view of the world was rejected by Aristotle. To Democritus 2,400 years ago, all things were composed of indivisible particles in constant motion. This concept might not have suited Aristotle, but it has been embraced by modern physicists.

Matter is now known to be composed of many kinds of entities that, despite Democritus, may be broken down into smaller ones. At various structural levels, however, it is frequently convenient to regard these divisible entities as being fundamental.

In condensed-matter physics, which includes, for example, the entire range of macroscopic liquids and solids, the basic units are considered to be atoms or molecules. At a finer level, the problem becomes the way in which electrons around nuclei form atoms. Turn up the magnification a notch, and neutrons and protons are seen moving inside the nucleus. At the greatest degree of magnification, represented by the highest energies achieved in the largest particle accelerator, the limit of description becomes the components of neutrons and protons: quarks and gluons.

At each structural level, the system is composed of identical interacting entities, and its properties are determined by the way the entities interact. The various interactions are governed by known force laws: electromagnetic laws for condensed-matter and atomic systems, strong force laws for particles, and both electromagnetic and strong force laws for nuclei. (Gravity and the weak force are insignificant at these structural levels.) To go from a force law or laws to a description of a particular system's properties, however, such knowledge is not enough; one also must confront a Medusa-like apparition.

Known in scientific argot as the many-body problem, this mathematical construct is really an assemblage of two-body problems, in that particles at every structural level—from electrons to the member bodies of the solar system—mainly interact as pairs. The minimum number of bodies needed to constitute a many-body problem is in fact two. Many physicists see nature in terms that increase this number by two, because these additional bodies complicate the problem in a way that irremediably changes its character.

An exact mathematical solution exists for a two-body system, and the two-body solution is the basis of an almost-exact solution for the three-body problem. For four bodies or more, "the essence of the many-body problem is that there is no exact solution," says Hobson Wildenthal of Drexel University. "One has to be satisfied with approximations."

Once a third body is introduced into a system, he explains, it disturbs the states of the original two bodies and invalidates the solution to their interaction. From then on, the bodies never let each other alone, thereby making impossible any exact two-body solution in the system.

If four is the minimum number of bodies in a respectable many-body problem, the maximum number to be handled depends on the structural level under study. In nuclear and atomic physics, the maximum number ranges from a dozen to hundreds of particles. In condensed-matter physics, it is 10^{23} particles, the quantity found in one cubic centimeter of condensed matter (the unit of measure used in that discipline). Particle physics more closely resembles condensed-matter physics than it does atomic and nuclear physics in its maximum number of bodies: Such bodies are pair-produced in abundance in the high-energy collisions used to study elementary particles.

Statistical techniques must be used when large numbers of particles interact. The system is sampled rather than precisely described. The answers can be nearly as accurate as those reached by exact treatment of the two-body problem, or just as accurate as those reached by the nearly exact treatment of the three-body problem. Thus, having to deal with 4 to 100 particles, as nuclear and atomic physicists do, is hardest of all. These scientists face the many-body problem containing the largest number of entities to keep track of.

Solutions to most of the many-body problems at any of the four structural levels of modern physics require enormous computational resources. (See "One physicist's simulation," accompanying this article.) This is because, with some exceptions in condensed-matter physics, they are quantum many-body problems and are governed by the laws of quantum mechanics.

These laws, by which microscopic systems behave, are often at odds with those that rule the everyday physical world. For example, the uncertainty principle (a seminal contribution of German physicist Werner Heisenberg) states that the position and the momentum of a quantum body cannot be specified simultaneously. From this uncertainty comes the idea of the dual nature of quantum bodies: sometimes particles, sometimes waves. Wave functions, which express the probability of finding quantum particles at a given place at a given time, must be used to describe the behavior of such bodies.

The equation devised by German physicist Erwin Schrödinger is the most commonly used formulation of many-body problems in quantum mechanics. The Schrödinger equation describes the evolution in time and space of a wave function. It is what mathematicians call a partial differential equation.

"The technical difference between partial differential equations and the ordinary differential equations used in classical mechanics is enormous," says Geoffrey Chester of Cornell University. "It makes the quantum many-body problem horribly hard to deal with."

Compounding the difficulty in many instances are extra terms that must be introduced into the Schrödinger equation to account for another quantum phenomenon elegantly described by the Pauli exclusion principle. Austrian-born

A body meets two. In an encounter of a double star with a single, simulation of the result defies computation.

Wildenthal. With grad student Xiangdong Ji: "There is no exact solution."

physicist Wolfgang Pauli added to the foundations of quantum mechanics by demonstrating that for a large class of particles called fermions, no two identical particles with the same orientation, either spin up or spin down, can occupy the same physical state in a system—can share a finite volume comparable to their own size.

(Fermions include many important particles, such as helium 3 atoms, electrons, protons, and quarks. In contrast, bosons make up a class of particles that do not obey the Pauli exclusion principle. They include helium 4 atoms, photons, and gluons.)

La Brecque was the author of "Fractal Symmetry" in Mosaic *Volume 16 Number 1.*

Algebraic or numerical

Such complexity did not becloud the solution to the first many-body problem addressed by physicists: the interactions among the planets in the solar system as they perturb dominant planet-sun interactions. The differential equations of motion devised by Isaac Newton and his successors to deal with that classical many-body problem are still the equations of choice for the latest many-body problems in astronomy. The dominant force is gravity, which is adequately described by classical physics, except for relativistic situations.

The concepts of celestial mechanics, fully developed by the school founded by Newton, culminated in a solution in the 1830s that was not improved upon for more than a century, until the development of computers made it possible to carry out the perturbative corrections of the planetary interactions to much greater accuracy.

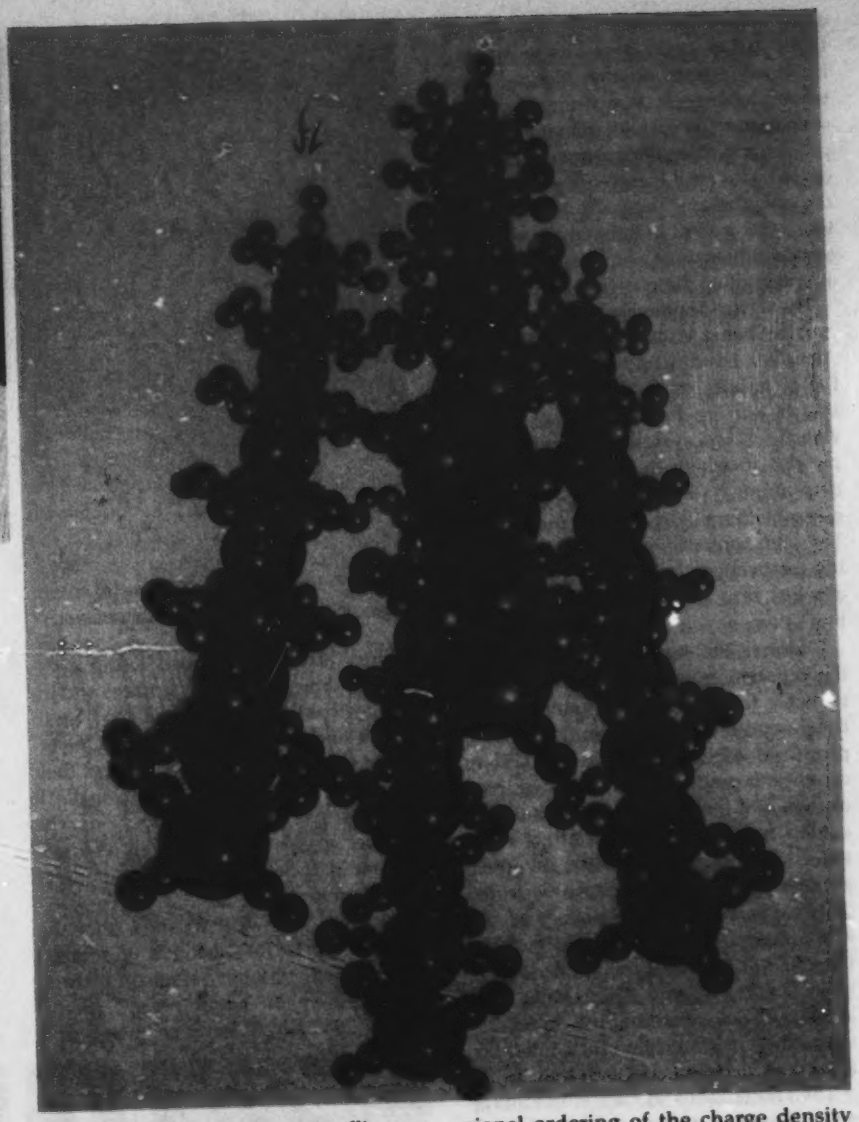
The basic approach used by the classical physicists was that of analytical mathematics, in which the detailed, point-by-point behavior of a system is expressed with a great economy in algebraic notation. All that is required in the way of computational tools is a piece of paper and a pencil.

The analytical approach was also used to attack a wide range of quantum many-body problems in the late 1920s and early 1930s, soon after the foundations of quantum mechanics were laid. Progress came abruptly to a halt, however, once the simplest problems had been solved. The mass of terms needed to obtain an algebraic solution to problems of greater physical complexity rapidly grew to a point where analytical methods no longer worked.

The mathematical techniques that have come to be used with these quantum many-body problems are numerical in nature: Each physical property is given an explicit value at every point in space or in time. A long list of numbers clearly lacks the elegance of an algebraic equation, but, in the words of Hobson Wildenthal, "as the additional complexity which truly characterizes nature at the most basic level becomes the cutting issue in science, you have to go to numerical techniques."

The computer is an absolute necessity for these techniques, having arrived fortuitously when the problems that were solvable analytically had by and large been solved. Analytical techniques still have a role in quantum many-body problems, however, and are themselves now being programmed for computers.

"Quite often, analytical work is done on idealized systems that contain the essence of a problem," says Douglas Scalapino of the University of California at Santa Barbara. "We reserve numerical



Scalapino and Krogmanns' salt. At critical temperatures, the charge-density-wave correlations become ordered.

simulation for real systems, particularly when they have large deviations from the ideal."

Scalapino is a condensed-matter physicist who has been studying, both analytically and numerically, quasi–one-dimensional materials. Including both organic solids and polymers, these materials are composed of ions, atoms with electric charges, arranged in stacks of chains with relatively large spaces between the chains.

Because of this structure, free electrons (electrons not tightly associated with a particular ion) can move easily along a chain, but they seldom hop from one chain to another.

With colleagues Jorge Hirsch, Robert Sugar, and Douglas Toussaint, Scalapino is using computer-simulation correlations to find discernible patterns that will indicate a system's specific state, for example charge-density-wave, magnetic, or superconducting. The wave-nature of quantum particles, electrons in Scalapino's work, plays an essential role in these calculations.

If a system is in a charge-density-wave state, the nearest neighbors of a site that has a charge will have no charge and the nearest neighbors of a site that has no charge will have a charge. In this way the density of charge is modulated along the chain. (See "Another Kind of Conductivity" by Edward Edelson, *Mosaic* Volume 15 Number 5.)

Recent studies show such a state in an organometallic material usually called Krogmann's salt, or KCP (shorthand for its chemical formula). Organometallics are of special interest because they contain large numbers of carbons, hydrogens, and nitrogens instead of being composed only of metal atoms, and they exhibit electrical properties.

"As we cool KCP and run it on the computer at lower and lower temperature, charge-density-wave correlations grow along the chain," says Scalapino. "Then, at a critical temperature, the system achieves a beautiful three-dimensional order; charges become ordered from one chain to another in the same charge/no-charge-charge pattern."

This simulation agrees with earlier X-ray scattering work in which charge-density-wave correlations along the chain are seen experimentally as a broad, diffuse structure, evidence of three-dimensional ordering of the charge density wave that comes into being as the temperature is lowered.

Two dimensions

Two-dimensional problems (one dimension more complex than quasi–one-dimensional problems) occur in surface physics, which has recently drawn increasing attention from the condensed-matter community. Surface physics consists of the study of phenomena that occur at or near surfaces (boundaries between a solid or liquid and a gas or vacuum) and interfaces (boundaries between a solid and either a liquid or another solid).

These phenomena are two-dimensional in that free electrons, restricted to a very narrow region of perhaps 25 to 50

angstroms in the direction perpendicular to the surface or interface, can move only in two directions on that plane. Many-body problems in two dimensions are even harder to deal with than many-body problems in three dimensions. This is because a powerful theorem that simplifies the calculations for periodic systems, those that extend indefinitely in all three spatial dimensions, cannot be used on aperiodic, two-dimensional systems.

The two-dimensional many-body problem being studied by Richard Wallis, Alexei Maradudin, and Adolfo Eguiluz of the University of California at Irvine is that of surface or near-surface atomic vibrations in crystalline materials—in metals—in which the phenomenon occurs normally at room temperature because of thermal energy in the crystal. The vibrations decrease in amplitude as the temperature is lowered, but even at absolute zero, a so-called zero-point motion occurs.

Thermal vibrations are of more than theoretical interest, for they have implications for technology. For example, electrical conductivity in inversion layers, produced when voltage is applied to common boundaries between semiconductors and insulators in Mosfets (metal oxide semiconductor field-effect transistors), is strongly affected by the interaction of the vibrations and free electrons. "We're interested in the motions of the atomic nuclei on the one hand and the motions of the electrons on the other," says Maradudin, "and of the interactions between the electrons and the nuclei, between the nuclei and the nuclei, and between the electrons and the electrons."

The interactions are all electromagnetic. However, because the particles are charged—the nuclei positively, the electrons negatively—the interactions between particles of the same charge repel and those between particles of opposite charge attract. Both kinds must be taken into account in the course of the treatment of the vibrations.

Pseudopotential

Before considering the vibrations, however, the Irvine physicists had to turn their attention to a more basic difficulty in surface physics: determining the position of the atoms in the system. The normally equidistant spacing between the atomic planes in a crystal breaks down near the surface; the precise relaxation of the surface atoms essential to any other surface calculations must be computed first.

Sugar. Configurations with probabilities.

One technique for predicting crystalline atomic positions at surfaces and interfaces has been developed by Marvin Cohen and Steven Louie of the University of California at Berkeley. All that their computer program requires to make it work is the atomic number, or the number of protons, positively charged nuclear particles, of the element under consideration.

Further, "We want to show that the method works for all kinds of solids," says Cohen. "Metals, semiconductors, insulators, even semimetals." (Semimetals are halfway between metals and semiconductors in electrical conductivity.) The ab initio pseudopotential method, as the technique is called, can predict a variety of physical properties at many levels of structure, says Cohen: molecular (2 to several atoms bound together); cluster (5 to 100 atoms); surface and interface; and the fully three-dimensional, or bulk.

To determine the structural properties of molecules, clusters, and solids, it is necessary to know the total energy of the system as a function of the atomic positions. Nature chooses the lowest energy configuration; hence the structure and stability of a solid are determined by interaction between the atoms.

It is not possible at present to calculate precisely the total energy of 10^{23} nuclei and electrons in one cubic centimeter of solid, all interacting with each other. However, the pseudopotential method can be used to obtain what Cohen and Louie call an "excellent approximation of the change in the total energy when the atomic positions are varied."

Some of the electrons in solids, the core electrons, are tightly bound to their atomic nuclei; other electrons, the valence or conduction electrons, run freely inside the crystal and really determine its properties. The pseudopotential method treats the core electrons around a given nucleus, together with the nucleus, as a single entity called the core, thereby markedly decreasing the number of interactions in the calculation; a valence electron interacts only with other valence electrons and with cores.

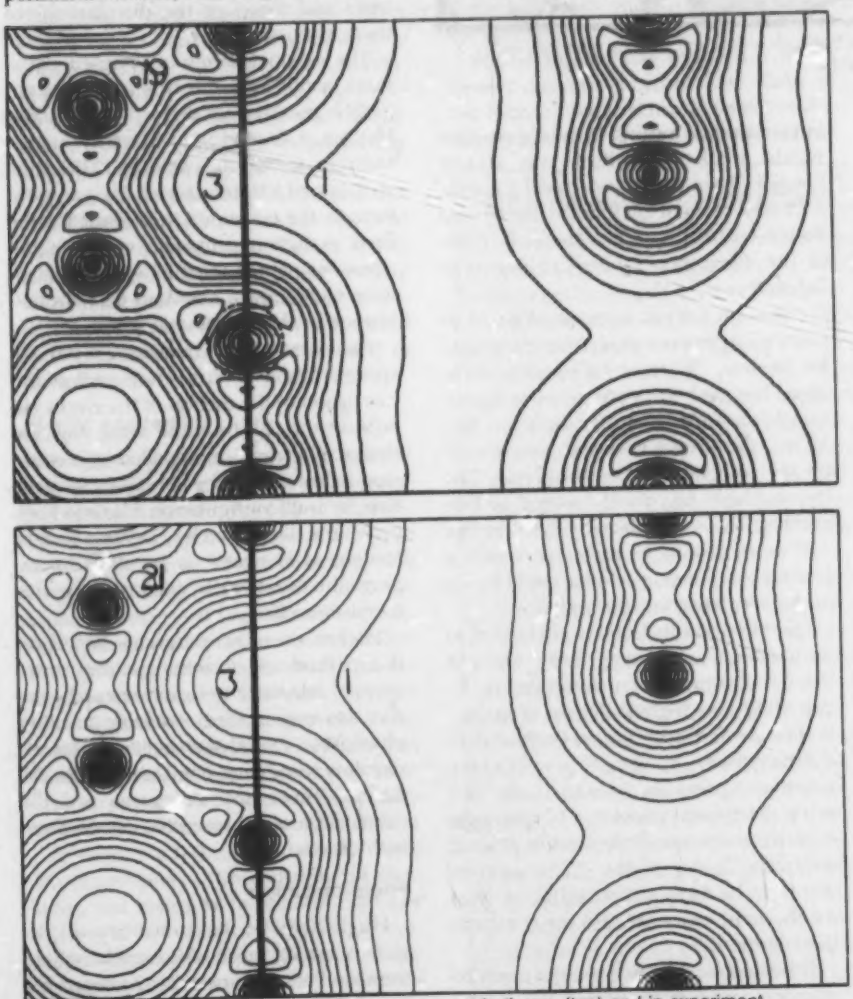
Cohen and Louie. All their programs need is the atomic number—the number of protons.

Another way the pseudopotential method simplifies the solution to the many-body problem is to effectively weaken the attractive force of the nuclei on the valence electrons. This weakening occurs because the valence electrons tend to avoid the core electrons as a consequence of the Pauli exclusion principle.

As the name indicates, the ab initio pseudopotential method, originated by Cohen and others and further developed by Cohen and Louie, operates from first principles, requiring no empirical input whatsoever. Cohen attempted to extend to surfaces and interfaces the empirical pseudopotential he had developed for bulk semiconductors. He found that the data developed by experimental work on these two-dimensional systems were not accurate enough to use with confidence in his calculations. The field needed a method that would be independent of experimental results, and he invented one.

Cohen uses the method to determine ground-state properties, such as structure, that characterize a crystal at its lowest energy configuration. Louie uses the method to study different kinds of properties—for instance, the intensity of electrons emitted by the crystal when light of different frequencies is shined on it. These kinds of properties require an excited-state many-body calculation.

The physical difference between a ground state and an excited state, Louie points out, varies with the nature of the crystal. In the ground state, valence or conduction electrons sit on the bonds that hold the atoms together in a semiconductor, stay near the atomic nuclei in an insulator, and are evenly distributed throughout the system in a metal. The behavior of conduction electrons in a metal does not change significantly when the metal is excited with small energies. When semiconductors and insulators are excited, however, their valence electrons are no longer localized, and they act like those in a metal.



Density maps. Contour plots in graphite: agreement in theory (top) and in experiment.

Principally, Cohen studies semiconductors like silicon and Louie looks at transition metals like platinum, but there are other, much more exotic, condensed-matter systems. Helium, for example, ordinarily thought of as a gas, becomes such a system as it is cooled; it first turns into what is called a quantum liquid, and then becomes a quantum solid.

Quantum liquid

"The low mass of helium atoms (only the hydrogen atom is lighter) and their very weak interactions with each other (all their electrons are tightly bound to the nucleus) make them quantum particles that can be described only by the Schrödinger equation," says Geoffrey Chester of Cornell University. With Malvin Kalos of New York University, Chester studies the two isotopes of helium, the fermion helium 3 and the boson helium 4, which has an additional neutron in its nucleus.

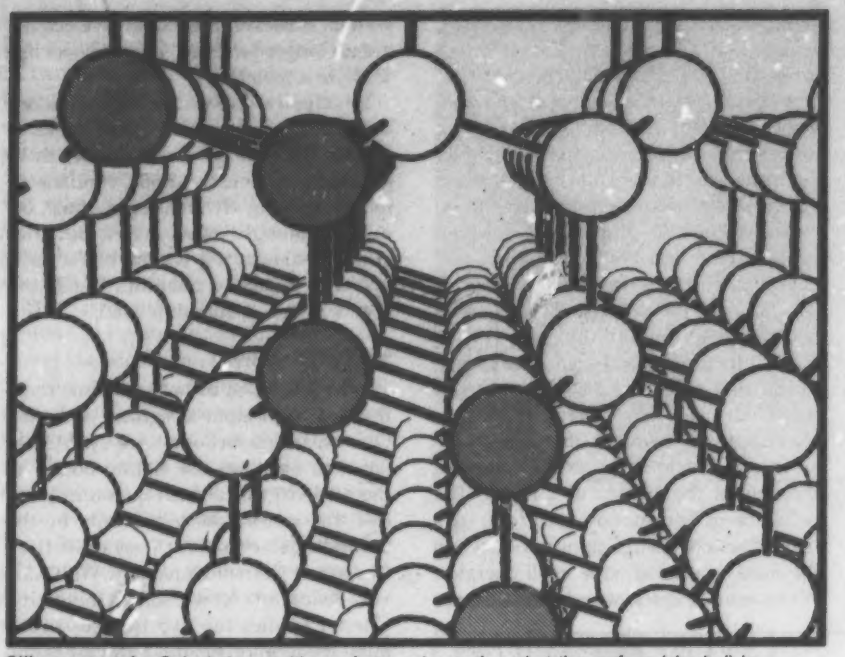
Despite the fact that one isotope obeys the Pauli exclusion principle and one does not, both have the same kind of extraordinary properties. As the gas phase is cooled, the helium atoms condense into a liquid that at a very low temperature becomes a superfluid. It loses frictional properties and flows through a pipe without any apparent resistance.

At an even lower temperature, superfluid liquid helium freezes and becomes a solid, but one unlike most other solids. The atoms of conventional solids make only small motions about their regular positions, which are quite close to each other. In contrast, the helium atoms are spaced very far apart and make large motions about their regular positions.

Chester and Kalos's computer simulations of the helium isotopes are so far in agreement with the extensive experimental observations of the isotopes. These results increase the researchers' confidence not only in the particular algorithm they use but, more importantly, in their basic computational technique, Quantum Monte Carlo.

Quantum Monte Carlo

Monte Carlo, developed for classical physics in the late 1940s, is statistical. It is a technique for studying the behavior of systems that have so many particles that a precise, deterministic approach is out of the question. The technique is not prac-



Silicon crystals. Calculating positions of atoms beneath and at the surface (shaded) in the crystalline arrays that make up semiconductors.

tical itself, however, without the use of another method, importance sampling.

"If different configurations of the particles were simply picked at random," explains Robert Sugar of the University of California at Santa Barbara, "almost all of the time they would be unimportant ones, and the process of taking averages of physical quantities would take enormous amounts of computer time. Importance sampling makes averaging efficient by picking configurations with a probability proportional to their importance."

Unlike the importance function in classical systems, which is known a priori, the importance function in quantum systems must be developed as the Monte Carlo calculations proceed. Finding a method for doing this proved extremely difficult, and it was not until the 1960s that one was invented by Malvin Kalos, after he had worked on the problem for about ten years.

Kalos's technique, however, was restricted to bosons, which are much simpler particles than fermions. Because they obey the Pauli exclusion principle, fermions can change sign when they exchange positions, as they do in solutions to the Schrödinger equation. This apparently innocuous quality greatly complicates their mathematical treatment.

Fermions are considered the more important class of quantum particles, so an importance-sampling technique that would work with them was clearly needed. Berni Alder and David Ceperly, a former student of Kalos's, developed such a technique back in the early 1980s at the Lawrence Livermore National Laboratory.

Although not yet as practical a tool as Kalos's importance-sampling technique for bosons, Alder and Ceperly's technique has made it possible to use Quantum Monte Carlo successfully on fermions. Geoffrey Chester is encouraged by the results of the simulations. "The two methods are closely related to one another," he says. "Intuitively, it seems that we're coming to the stage where a general computation scheme could be set up for condensed-matter systems."

Quantum Monte Carlo is not limited to such systems, however. Berni Alder is using the technique on atoms and molecules, the next highest degree of magnification, where electrons are the bodies in the many-body problem. "We want to see whether Quantum Monte Carlo can solve the typical problems of quantum chemists more accurately than traditional methods," says Alder. "The answer seems to be fairly encouraging, at least for the light elements and for combinations of them."

Other atomic scientists are not ready to abandon tradition in solving the many-body problem. Charlotte Froese Fischer of Vanderbilt University, for example, has updated the venerable Hartree-Fock method to predict atomic properties of bound-state systems, those in which all electrons are tightly bound in their orbits around nuclei.

Old but good

Unlike Quantum Monte Carlo, Hartree-Fock treats each electron as if it were moving independently of the other electrons. They share a characteristic, however, in that the distributions of the electrons, the probability of finding them at various distances from the nucleus, are interrelated.

An electron responds to a field in which the attractive force of the nucleus is screened by a cloud made of all the other electrons. This field determines the electron's distribution, which helps determine the field affecting the other electrons and therefore the distributions of those electrons.

The original series of coupled equations resulting from this independent particle model was derived by Douglas Hartree, a Cambridge University mathematician, in the late 1920s. In 1933 a Russian named Vladimir Fock added an extra term to the equations to account for the Pauli exclusion principle, enabling the equations named for the two mathematicians to become a standard first approximation of the real atomic world.

The Hartree-Fock equations are only an approximation because they neglect the correlation in the motion of the electrons; because they all have the same negative charge, electrons are impelled by electromagnetism to avoid each other. A variation, a multiconfiguration Hartree-Fock procedure developed by Fischer, a former student of Hartree's, is an effort to improve the equations by incorporating this correlation effect.

Fischer uses the programs based on this method in determining atomic properties of relevance to fusion research, but she also makes them available to other physicists. "I've always believed that research is something one shares," she declares. "Since my programs are the culmination of my research, I share them with others."

Photoionization

Hugh Kelly of the University of Virginia is among those who have benefited from Fischer's openness. He has used her programs as a starting point for his stud-

ies of photoionization. The free electrons in this process are influenced by a bound system that can be described by use of Fischer's code.

In photoionization, an atom ejects an electron after being bombarded by one or more photons. The incoming particles impart their energy to the electron, which does not have enough energy of its own to overcome its attraction to the nucleus; if the imparted energy is sufficient, the bound electron can free itself, leaving the formerly neutral atom in an ionized, positively charged state.

Because it occurs whenever matter is hit by radiation of the right frequency range, photoionization is studied in different systems to understand the transport of radiation; that is, exactly how radiation is absorbed at different frequencies. Kelly has computed photoionization characteristics for many atoms, including some that are high on the periodic table. "One might think the many-body effect would be unimportant in the calculations," he says, "but the process is very much affected in many cases by the presence of the other electrons. We usually have to treat all the electrons interacting simultaneously among themselves."

An inner electron may be the one that absorbs the photon, for instance, and goes to an excited state. Before it can leave the atom, though, it may collide with another electron and transfer its energy to that struck electron, which would then become the one that is enabled to leave the atom.

Elastic scattering

A related problem is electron scattering, in which the atom is hit by an external electron rather than by a photon. There are two types: elastic scattering, in which the electron, after interacting with the atom, leaves with exactly the same amount of energy it had when it came in, and inelastic scattering, in which the electron gives up some of its energy to the atom before leaving. The atom emerges from an inelastic scattering event in an electronically excited state. Because it can rotate and vibrate, a molecule involved in such an event may be excited rotationally and vibrationally, as well as electronically.

Inelastic scattering is usually the more important process in collisions of electrons and atoms or molecules, says Lee Collins of Los Alamos National Laboratory, who, with his colleague Barry Schneider, has been simulating such collisions. Nevertheless, notes Collins, more attention has been paid so far to elastic scattering, particularly in molecules, because of its relative simplicity.

Electron scattering occurs in lasers as the driving force that excites a population of molecules to the energy state from which the coherent beam of photons emerges. It also acts in the controlled nuclear fusion devices called Tokomaks. Colliding against the metal walls of the Tokomaks, energetic electrons create undesirable impurities in the plasma confined by the machine's walls. Collins and Schneider's methods can thus be used as a diagnostic tool.

Scattering studies also help diagnose larger-scale, natural occurrences. For example, processes ionizing the gas molecules in the ionosphere free electrons that hit other molecules there. Excited molecular states resulting from these collisions of free electrons produce radiation, which can be analyzed with respect to scattering to determine conditions within the ionosphere.

Radiation that has traveled a much greater distance concerns Kate Kirby of the Harvard-Smithsonian Center for Astrophysics, who studies atomic and molecular systems that are astronomically interesting. These include carbon, nitrogen, and oxygen, as well as hydrogen and helium.

Fischer. "Something one shares."

"Most of what we know about astrophysical objects comes from detailed studies of the light emitted or absorbed by their constituent atoms and mole-

One physicist's simulation

To solve the many-body problem for quasi-one-dimensional materials such as the organometallic material called Krogmann's salt, or KCP, condensed-matter physicists Douglas Scalapino and Robert Sugar of the University of California at Santa Barbara must study 10 × 10 arrays of chains, each of which is 30 lattice sites long. The size of this three-dimensional system, a miniature crystal of the material, permits the results of the computer simulation to be extrapolated with confidence to the material in bulk.

Unlike classical physics systems, quantum-mechanical systems cannot be isolated at one point in time; instead, they must be thought of as oscillating over many such points, a particular position in space being associated with each point. The number of time points required for an adequate description of KCP is 60, so the total number of lattice sites that must be considered is 10 × 10 × 30 × 60, or 180,000.

Each site, moreover, has 40 variables, four possible electron configurations (no electrons; one electron, spin up; one electron, spin down; and two electrons, one spin up and one spin down) multiplied by ten possible values for the site's spatial position. Before the properties of the system can be computed, however, 100,000 of its $40^{180,000}$ possible configurations must be generated with the Quantum Monte Carlo technique—in probabilities proportional to their actual occurrence—and considered for each variable for each site.

Every change of configuration requires about 100 additions or multiplications, called floating-point operations, or flops. The total number of flops for a single run on KCP, therefore, would be 180,000 × 40 × 10 × 100,000 × 100, or 7.2×10^{14}. That would take the advanced Star Technologies ST-100 array processor used by Scalapino and Sugar, running nonstop at its maximum speed of 10^8 flops per second, a total of 10^6 seconds, or about 12 days.

Because many runs are needed for a full system examination, the numerical simulation of KCP could easily require several months or more on a state-of-the-art computer. It would be inconceivable without one. •

Kirby. Collisions of atoms with photons help interpretation of astronomical observations.

cules," says Kirby. Knowledge of "the properties of atoms and molecules and the response of the atoms and molecules to collisions with photons is needed to understand the processes occurring in such objects and to interpret astronomical observations made with telescopes."

Her many-body problems involve the atoms and molecules in the atmosphere around a star, for example, which particles absorb light emitted from the burning of matter in the stellar interior, make a transition to a higher energy level, and then emit light themselves. Other astrophysical sources of light are the atoms and molecules in the highly shocked region where a supernova has exploded.

Astrophysics

By splitting the radiation coming from an astrophysical object into its spectral frequencies, astronomers can identify the atoms or molecules in the vicinity of the object. The different species all absorb and emit radiation at characteristic frequencies. A dark spectral line in a light spectrum indicates the presence of a particular atom or molecule that is absorbing radiation; a very bright line indicates the presence of an atom or molecule that is emitting radiation.

A focal point of Kirby's current research is the carbon monoxide molecule, important astrophysically because it is found in the atmospheres of the sun, stars cooler than the sun, planets, and abundantly in the interstellar medium. Carbon monoxide is also used to map the mass distribution of distant galaxies.

The molecule can be destroyed by photodissociation, a process in which the bonds between atoms break as a result of the absorption of photons of certain frequencies. In the case of a carbon monoxide molecule, the fragments left behind are, of course, a carbon atom and an oxygen atom. "In considering formation and destruction mechanisms for astrophysical molecules, photodissociation may be a significant pathway for destruction of molecules," says Kirby. "I'm trying to compute an accurate dissociation rate for carbon monoxide."

Other astrophysical questions have been answered almost inadvertently by nuclear physicists, whose many-body simulations involve nucleons, the positively charged protons and uncharged neutrons that interact in the atomic nucleus. Because nucleons are packed together very closely, when Vijay Pandharipande of the University of Illinois studied the manner in which they interact, the results of his studies also became relevant for matter in neutron stars. (According to current theory, neutron stars are the remnants of supernova explosions of medium-sized stars. Such a star has a core of elementary particles surrounded by a fluid composed primarily of neutrons, the density of which is far greater than that of ordinary nuclear matter.)

Although the bodies in a many-body problem interact principally as pairs, they can also have very weak three-body interactions, which involve every possible combination of three bodies in the system. The strength of three-body interactions can be determined by subtracting from the observed properties of a three-body system the interactions of the three pairs of bodies in the system. (No such empirical evidence exists for significant four-way interactions between every combination of four bodies in the system.)

Pandharipande's simulations of interactions between many bodies at the

Kogut. Quantum chromodynamics manipulated so as to simulate quark confinement.

Hartree-Fock wave functions. Each electron of an argon atom is treated in these computations as if it were moving independently.

high densities that exist in neutron stars and atomic nuclei showed that the three-body interactions, though weak, are quite important. They explain why the binding energies and densities of nuclei are not greater than those observed, and that they stabilize neutron stars enough to increase their maximum radii and masses.

A nuclear model

One of the most intriguing many-body problems faced by nuclear physicists recalls the description of the planets in orbit around the sun. The planetary system had proved to be an obvious model for the atom with its electrons orbiting around a nucleus. Since the number of bodies in a nucleus is roughly the same as those in an atom, the planetary system seemed a possible model for the nucleus as well. However, problems seemed to abound: Because nucleons (unlike electrons and planets) are so tightly packed as to be considered in contact and also interact strongly with each other, it was hard to see how they could be moving smoothly in orbits; nucleons should bump each other out of orbit continuously.

The insight needed to resolve the quandary came when the nature of quantum mechanical particles such as nucleons was considered more deeply. The Pauli exclusion principle dictates that only those rare, violent collisions that bump a particle into unoccupied space can actually occur. The more probable, less violent collisions that would bump them into already occupied orbits are forbidden. Thus, the orbiting motion of nucleons is much more stable than it first appeared.

The geometrical picture of the nucleons that emerges is quite similar to that of electrons, which, as quantum-mechanical particles, are sets of concentric spheres, called shells, each allowed by the Pauli principle to contain a certain maximum number of particles. Particles in the innermost shell are in the lowest energy state, and those in the outer shells are in successively higher energy states.

Occasionally, there is a big jump in energy between groups of shells. If the gap between groups of shells is large, it is reasonable that the typical behavior of the system would depend only on the particles in the shells beyond the gap. The shells below the gap are uniformly filled with particles and effectively can be dropped from consideration in treatment of the many-body problem.

One thing distinguishes this shell model of the nucleus from the atomic shell model: The nucleus lacks a central object holding it together; that is, the nucleus lacks a nucleus. Instead, the average effect of the paired and three-body interactions of all the nucleons is equivalent to that of an attractive central force, with something left over.

This residual force, as it is called, is analogous to the perturbations from particle-particle interactions that occur in the solar system (planet-planet) and in the atom (electron-electron), but has a greater relative strength. It is almost as large as the average force pulling each individual nucleon to the hypothetical center. "The residual interactions actually govern the great mass of phenomena we observe in nuclei," says Wildenthal. "They determine the many different ways that nucleons can be distributed in the partially filled outer nuclear shells."

Wildenthal is working on what he calls a neoclassical shell model—neoclassical because it is conceptually the same as the nuclear shell model first proposed by Maria Goeppert Mayer in 1947. That model was a one-particle model; it dealt with filled inner shells and an outer shell with only one particle, which alone determined the properties of the nucleus.

"Mayer couldn't go on to the next level of complexity, several outer shells, because she didn't have available then the computational machinery to handle the interactions between the outer nucleons," says Wildenthal. "I'm using the computer to calculate and add up all of

Krypton and lanthanum. Nuclei containing many nucleons move toward each other, collide and fuse, then break apart.

the interactions between as many as 24 nucleons in three shells. These calculations represent the nuclei of the 12 elements from oxygen to calcium in the periodic table."

His approach is to use experimentally observed spectra of these complex nuclei as a standard by which to obtain the correct values of the residual interactions, essentially unknowable except in qualitative outline. Since each change in the residual interaction changes the nature of the system itself, the corrections must be done in a circular or iterative fashion. The calculations are run over and over on the computer until they converge, until the last set of corrections is negligibly small.

Simulating collisions

Hobson Wildenthal's calculations are undoubtedly of interest to Steven Koonin of the California Institute of Technology. The shell model of the nucleus is the basis of Koonin's simulations of low-energy collisions between heavy nuclei, which are nuclei with as many as 240 nucleons apiece. (More than 2,000 nuclei are now known for the isotopes of the 109 elements; each nucleus has from one to 109 protons, and each has a number of neutrons that is sometimes equal to but often greater than the number of protons.)

Koonin's simulations are a theoretician's version of the collisions used by experimentalists since Ernest Rutherford in 1910. Beams of nuclei are shot at a nuclear target in a particle accelerator in the hope that the nuclei will form a combined entity. If they do, the entity is invariably an unbound system, one that quickly—in about 10^{-17} second—breaks into pieces whose properties can be observed and studied.

Because heavy nuclei can be made to collide only at relatively low energies in the current generation of accelerators, the simulations so far have been held to those conditions. The study of heavy nuclei is aimed at understanding of the behavior of nuclear matter in bulk. "As nuclei get larger, their surfaces get smaller in relation to their total volumes," says Koonin, "so they better represent a bulk system."

Moreover, an island of superheavy nuclei has been predicted whose inhabitants may be long-lived enough to be observed. Chemists hope to make these superheavies by the collision of two large nuclei, through either transferring a piece of one to the whole of the other or getting the two to stick together and emit several lighter particles.

Not all the collisions Koonin simulates are head-on, or central, collisions. The existence of peripheral collisions, in which the two nuclei only sideswipe each other, perhaps exchanging a few nucleons and a little energy, has long been known. About ten years ago, an intermediate type, the deep inelastic collision, was discovered.

In a deep inelastic collision, as in a central collision, there is a large exchange of mass and energy and a large loss of energy from motion into heat. The combined nucleus that comes out of a deep inelastic event, however, appears to retain some memory of its original components. It breaks into two fragments that resemble those components. The result is about the same as that of a peripheral collision.

The result of a central collision, Koonin says, depends on two properties of the compound nucleus. If its energy is not too high and spin not too great, the nucleus will decay by emitting photons, neutrons, protons, and alpha particles (each of which consists of two protons and two neutrons). The single large piece left behind may be a rather exotic nucleus with an unusual number of neutrons and protons.

If the energy of the compound nucleus is high or the spin is great,

however, the nucleus will split into two or three pieces—extended and distorted objects that fly apart as rapidly as several times 10^{-23} second. These pieces lose some of their energy by emitting photons, and they eventually emerge as nuclei themselves.

All of this can be tracked in the simulations. "What we have typically done," says Koonin, "is to observe the locations of the nucleons during the collision, how the shape of the compound nucleus changes with time, and then what happens to it after it has broken into fragments."

Simulating quark confinement

Nuclear properties become apparent when heavy nuclei are collided at low energies; similarly, the high-energy collisions of protons, which are, along with neutrons, the basic constituents of nuclei, point toward the protons' basic constituents, quarks and gluons, as far as physics can tell. (See "Physicists' Symmetries" by Margaret L. Silbar, *Mosaic* Volume 16 Number 1.)

Elementary-particle physicists are trying to confirm that the proton is indeed made of three quarks in a field of massless particles called gluons. As the carriers of the strong force holding the quarks together, gluons are the cornerstone of the theory known as quantum chromodynamics. This is the underlying theory of nuclear physics that scientists hope will eventually provide an understanding of the fundamental interaction between neutrons and protons. This is a function of electromagnetism as well as the strong nuclear force, since protons repel each other.

Quantum chromodynamics is the quantum field theory that describes the strong nuclear force as the older such theory, quantum electrodynamics, describes electromagnetism. Instead of dealing with electrons, positrons, and photons (as electromagnetism does), however, this theory deals with quarks, antiquarks, and gluons. The charge that affects the gluons is not electric but—in the argot of particle physicists—color, hence the prefix chromo. There are three color charges: red, blue, and yellow. The net color of any particle that can be isolated physically must be neutral. A color-neutral particle can be formed from a red quark, a blue quark, and a yellow quark or, say, from a red quark and a red antiquark.

Color is believed to be the source of flux tubes that form, experiment indicates, between these two kinds of quark assemblages. The flux tubes, in turn, are the source of an energy potential that grows with the distance between the quarks and confines the quarks irreversibly in such assemblages; a quark has never been detected all by itself. (See "Quarks in Nuclei" by Margaret L. Silbar, *Mosaic* Volume 16 Number 2.)

John Kogut of the University of Illinois at Urbana-Champaign is currently simulating quark confinement through the use of lattice gauge theory, a recent formulation of quantum chromodynamics in numerical terms that can be programmed on a computer. The lattice in the theory is a fine grid on which an infinity of possible space-time values is approximated by a set of points.

"Right now, none of the work on lattice gauge theory is really decisive," says Kogut. "An early simulation of confinement that had a quark and an antiquark sitting in a gluon field did the gluon field accurately but entirely left out any quarks between the quark and the antiquark. We're trying to put the quarks in the quark-gluon plasma on the same footing as the gluons."

Kogut must turn for these studies to the techniques of statistical mechanics used by condensed-matter physicists. Solving numerically the equations of quantum chromodynamics is a problem very similar to the problems in statistical mechanics that have been faced in the past. In this way the smallest and largest manifestations of the quantum many-body problem become one. ●

TAUNTING THE

by Norman Metzger

In the 1950s, neurophysiologists David Hubel and Torsten Wiesel discovered that specific neurons fired in response to specific visual elements, such as angled lines and bars. That work led to a natural conclusion: Vision depends on "feature detectors" that locate vertical, horizontal, or angled lines in an image. The brain then compares those features with information already resident and composes a scene.

The next step was for scientists working in the then young and overconfident world of artificial intelligence to translate the biological result into a digital one by equating images stored in computer memory with electronically sensed features. The immediate difficulty, of course, was that the world is more than a mosaic of bars and lines. The heroic efforts of the 1950s and 1960s to simulate vision by stuffing computer memories with digital forms of drawings, photographs, and shapes, from toy blocks to teddy bears, apparently came to nought before the reality of vision's complexity.

The word *apparently* is important, for despite its shortcomings, the work did have important results: (1) it led to useful applications in vision problems in which lines and bars were major elements and an ability to see the world in three dimensions was not (see "Computer vision" accompanying this article), (2) it suggested other avenues of attacking the vision problem, and (3) paradoxically, by its failure and by its revelation of how complex vision really is, it forged a bond between neurobiology and computing.

That bond has other roots: in links between the view of the neuron as an all-or-none firing device and the design by John von Neumann of the serial logic computer, and in the timeless comparisons of computer and brain, one a digital device and the other an analogue, one a serial processor and the other a parallel processor, one limited in the interconnections of its chips and the other connected in apparently infinite and not necessarily immutable ways.

Today, the working bond between computing and neurobiology serves in at least three ways. As it does for many fields, the computer gives neurobiologists an entry into otherwise hidden worlds. For example, computers are used to reconstruct intricate neural anatomies, in effect serving, as Cyrus Levinthal of Columbia University puts it, as "a 3-d notebook." In the same vein, computers have been used to model processes that are, by virtue of their scale, beyond direct visualization—in work, for example, by Rodolfo Llinas of the New York University School of Medicine on the governors of the synaptic communication between nerve cells.

Computers can also be used as amiable technicians; for example, computers make it practical for a number of laboratories in the United States and other countries to plan multineuronal experiments whereby a pincushion of electrodes implanted in an animal's brain signals the responses—whether causal or corollary being a point of contention—of several neurons to a stimulus. Not surprisingly, these kinds of experiments yield, as Llinas puts it, "an absolute avalanche of data."

Finally, the computer can suggest rather than just abet neurobiological experiments. The example is again vision; current work on algorithms to endow computers with stereovision in turn suggests biological experiments. Such work meets on the terrain of information processing, the terrain that says look at what the task is and ignore, for the moment, the "hardware" that does it, regardless of whether that hardware is biological or electronic. Decouple function from form.

A field in ferment

Before adding detail to these three computational streams in neurobiology—the computer as an instrument, as a technician, and as driver of experiments—some brief words about

Neural hierarchy. Almost perfect images from a skin nerve fiber in a monkey's finger.

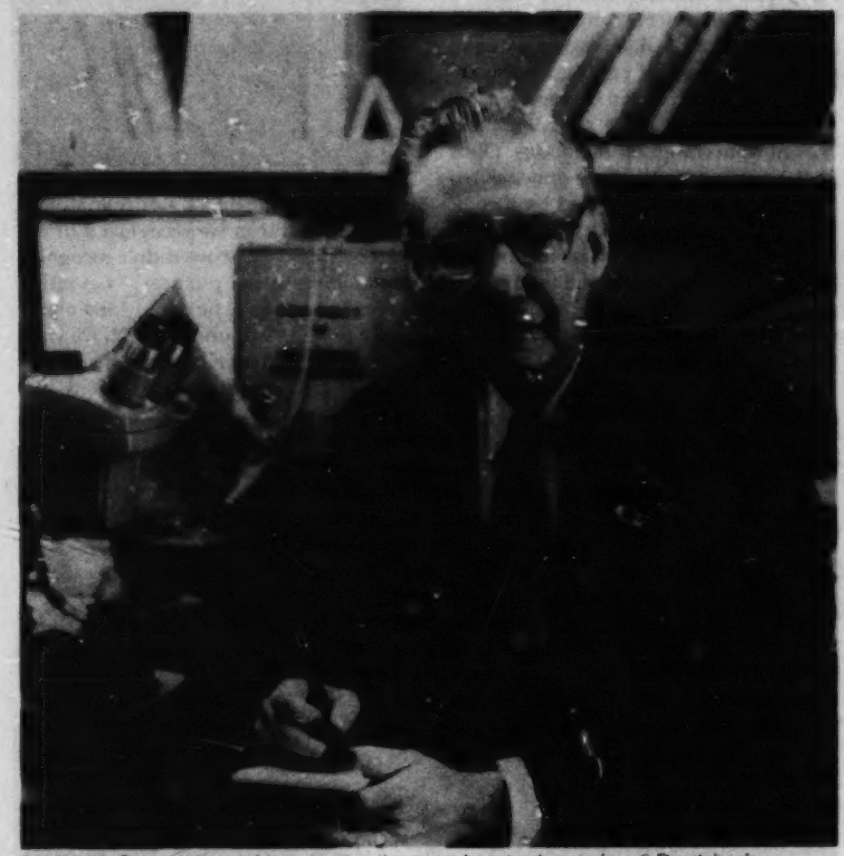
COMPUTER

The chimerical similarities between neurobiology and computing continue to dissolve on approach. But the fields do support each other in many ways.

the field itself, simply to moot glib comparisons with other fields in which the use of computers is considerably more robust. Part of the problem lies in the experimental difficulties. As George Gerstein of the University of Pennsylvania observes, "You need large teams to do this effectively. We're getting away from one man, one lab, where you go into the setup in the morning and you make your preparation and you stick in your electrodes and you photograph the face of your oscilloscope ten times and at five o'clock you go home. It doesn't work that way anymore."

Also, or perhaps because of these logistical difficulties, there is a limited population of researchers in any particular research niche. Tomaso Poggio of the artificial intelligence laboratory at the Massachusetts Institute of Technology says that there are only one-and-a-half people in the entire world working on the physiology of stereovision. One consequence, Gerstein says, is that "there is no good way to trade data or people or experience, and so the strengths of each laboratory cannot be merged." In fields such as molecular biology it is relatively easy to replicate an experiment; in neurobiology, replication becomes much more difficult, given the resources and time required.

Then there are the conceptual difficulties, made clear in a sprinkling of statements. Vernon Mountcastle of Johns Hopkins University notes, "One of the problems with brain science is that there isn't any central theory. Most neurobiologists are shaking the tree and catching what falls." In Levinthal's opinion, "The real difficulty is that as neurobiologists we don't know how to define the problem. It's not really fair to say, 'I've got all this data, now do something with it.' We're at such an elemental stage, we don't know what questions to ask." Poggio adds, "Real progress will come when you have theories or models you can test. It's not enough to set up correlation coefficients."



Levinthal. Computers used in reconstructing neural anatomies, as in a 3-D notebook.

The lack of central theory devolves into inconsistent views on the role of computers in neurobiology. Levinthal sees "no great need for supercomputers." Gerstein doesn't know "to what extent the problems we have in the laboratory would be amenable to parallel computation." In contrast, the algorithms for vision, in requiring the summation of data, such as shading, texture, brightness, and motion, on an infinite number of points in an image, force repetitive and identical calculations of the sort suited to parallel architectures. Indeed, a stereoalgorithm developed by Poggio and his colleagues is already running on a machine built by Thinking Machines, Incorporated, with an array of 128 × 128 processors.

Cartos

The brain taunts the computer, and its makers, with its wiring processes. The interconnections among the brain's "processors," neurons, are complex and multidimensional. That neural grid is surely essential to the brain's ability to compute at an astonishing rate, to recognize a familiar face in milliseconds. The begged question is, of course, how are neurons wired together correctly. As Levinthal notes, "In the primate, you've got a few million fibers going from the eye to the brain. How do they know

where to go? It seemed to me that one of the obvious things that had to be done was to look at the patterns as they grew." (See "Targets in the Brain" by Joann Rodgers and "Molecules that Guide Behavior" by Joseph Alper, *Mosaic* Volume 15 Number 5.)

The result was *Cartos:* Computer-Aided Reconstruction by Tracing of Sections. *Cartos* allows light or electron micrographs of serially sectioned brain or neural tissues to be converted—manually at first and, more recently, semi-automatically, using contour-following algorithms—into digital information, arrays of numbers, to be stored and retrieved from computer memories. Once in the computer, the three-dimensional images can be reconstructed on a computer screen and played with: turned, sectioned, dissected. By using models adopted from the fields of fluid dynamics or electrical transmission, *Cartos* reconstructions can be used to predict, for example, impulse propagation in a nerve fiber.

Levinthal makes "movies" of his reconstructions by aligning contiguous micrographs and then photographing each one with a pin-registered, 35-millimeter instrumentation camera. The product is a series of aligned photographs, a 35-millimeter movie. *Cartos* was made generally available to the research community through a loaner program, and it, as well as other systems for reconstructing neural tissue via computer, is now a common tool in neurobiological research.

A unique view

The reconstructions, by virtue of being three-dimensional and manipulable, offer a uniquely prismatic view of the tangled world of neuroanatomy. "There are things we recognized in the reconstruction that we didn't recognize by looking at the pictures," Levinthal says. For example, he adds, "one of our early projects was one in which we were looking at the development of optic fibers in a small water flea, *Daphnia*. Because we did the reconstruction, we know that this fiber was connected to a cell body a mile away. We simply wouldn't have seen the phenomenon without the reconstruction."

The tenor of the *Daphnia* work has been extended to other creatures: leeches and, most recently, zebrafish. Other investigators have used *Cartos* for studies of monkey cortex cells, shrimp muscle fibers, moth antennas, and the ultrastructure of bacteria. Recently, Levinthal and his colleagues have joined reconstructions to molecular biology, combining *Cartos*-developed insights

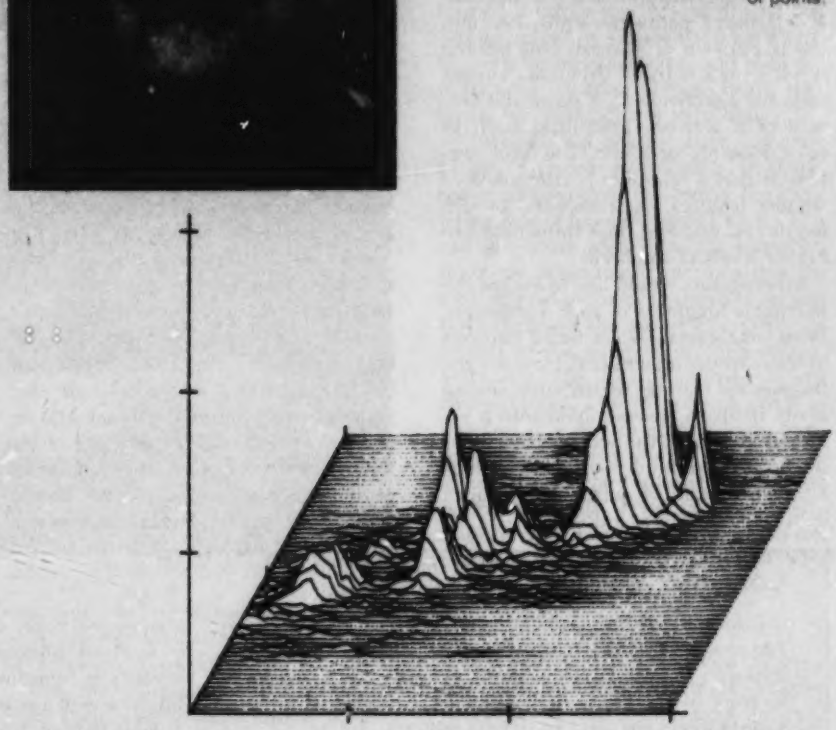
Poggio. With rabbit and digital rabbit. More regularity than people think.

into neural structure at different developmental stages with techniques, such as *in situ* hybridization, that show which genes are turned on and when.

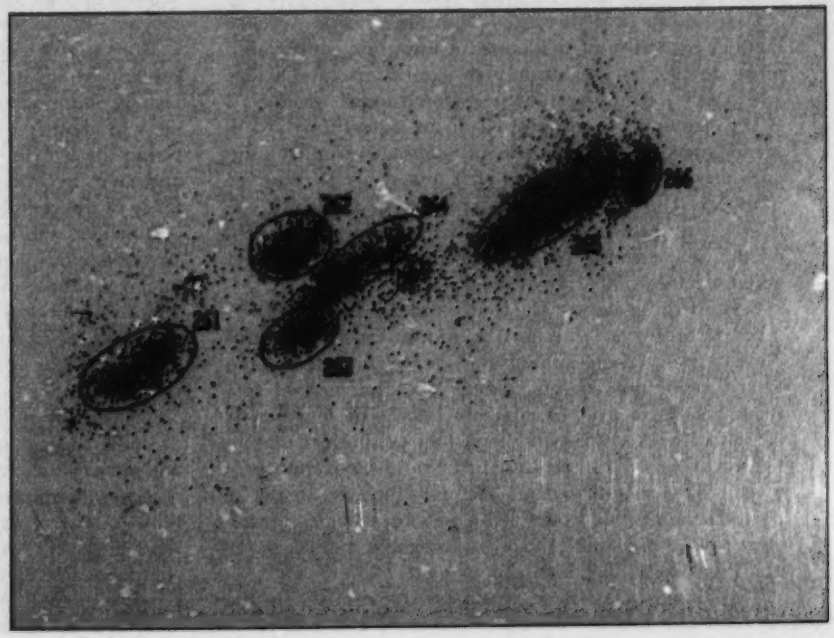
The problem with *Cartos* and other reconstruction methods is that entering information stored in micrographs into computers is largely a manual and hence an onerous chore, especially since one experiment may need 500 to 1,000 electron microscope sections to record 50 to 100 microns of tissues in sufficient anatomical detail. Of course, the computer should be able to do it. However, Levinthal points out, "you frequently do not get a membrane that is uniformly dark all the way around and without gaps. So, if you use a simple algorithm—a contour-following algorithm—you fall out through the gaps. If there's a mitochondrion in the cell, the program will frequently invaginate and follow the mitochondrion around."

One approach to sharpening up what Levinthal calls this "stupid but fast" contour-following algorithm is to complement it with an algorithm suggested by Peter Selfridge, of AT&T Bell Laboratories. Very simply put, the algorithm, now being developed by Robert Schehr at Columbia, looks for edges, sharp changes in image intensities. If it finds an edge, it retraces its steps, constrained by well-defined subalgorithms to search for a closer intensity match.

Gerstein. Ticking particles come together.



Spike-shaped sorters. Output of points from action potentials of neurons (bottom), and 3-D plot of the output with peak heights correlating to number of points.



The enchanted loom

The medium—neurons, their synaptic connections, and the like—is not, however, the message. In what form is information, a stimulus, layered upon a neural network? It is an old question. Indeed, one of the most poetic answers, although based on no evidence, was provided some 40 years ago in a description of the waking brain by the British physiologist and Nobelist, Sir Charles Sherrington:

> The brain is waking and with it the mind is returning. It is as if the Milky Way entered upon some cosmic dance. Swiftly the head-mass becomes an enchanted loom where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one; a shifting harmony of sub-patterns.

Sherrington's vision of the mind has endured. The difficulty, says George Gerstein, "is that we've known nothing

about assemblies until now." *Now* refers to the efforts of several laboratories—in Israel, the Netherlands, and the United States, and specifically in the laboratories of Gerstein, Mountcastle, and Llinas—to record the responses of many neurons to a single stimulus.

To date, recordings have commonly come from a single electrode embedded near a single neuron that is part of whatever sensory pathway—sight, hearing, touch, pain—is of interest. One finding of such work is that individual neurons respond selectively to a particular feature of a sensory stimulus, such as color, location, or shape. The basic idea now is that a stimulus is composed of simpler features, each encoded by different neurons and then transmitted to higher levels of the brain.

Investigators would like to escape the inevitable fogginess of such a situation. What amalgamations of firing neurons in toto capture a stimulus? How are the fragmented features, distributed among many neurons, reassembled into a coherent picture of the world? How does the brain "compute" a stimulus such as a touch, a whiff of perfume, a flash of light, the barking of a dog?

One working hypothesis is that it is not the structure of the neurons or their synaptic connections that matter but rather what they are doing in relation to one another. Vernon Mountcastle explains, "The information resides in the temporal relation of the activity of the different fibers. Knowing about synaptic transmission doesn't tell us anything about how the brain operates. The important business is done in the population activity." Or, as Llinas puts it, "If you want to understand the properties of these networks, you cannot understand them at the single cell level because a lot of information is present simultaneously at the wave front. You have to be able to test the population at a particular time."

Another working hypothesis is that the representation of a stimulus changes with its level in the neural hierarchy. Kenneth Johnson, Mountcastle's colleague at Johns Hopkins, stepped an embossed letter *a* across a single skin nerve fiber in a monkey's finger, and recovered an almost perfect image of the letter in the recordings. In other words, as Johnson points out, "at lower levels, such as in skin fibers, neural representation is essentially isomorphic—the neural patterns, properly massaged, will reveal the letter *a*." However, at the higher levels of the cortex, representations become less isomorphic and less obvious. "The cortical neurons are not, in any obvious way, firing in response to the shape of the letter," notes Johnson. Therefore, says Mountcastle, "the experiment has now shifted to the part of the sensory cortex to which the fiber in the monkey's skin projects." But, he adds, "since timing gets sloppy on the way up from skin to cortex, to really do the experiment right, we have to obtain readings simultaneously from a significant number of cells."

Multineuronal

Both Llinas and Gerstein are already at the level of multineuronal recordings, with Llinas working in the cerebellum and Gerstein with the auditory parts of the sensory cortex. Llinas, for example, is "recording simultaneously from 32 cells, using 32 probes." The work carries a multitude of burdens. One is skepticism by some that it will produce anything useful. A second is repeatability, since it is now impossible, except by

Computer vision

Computer vision is already a thriving industry. Some 100 companies make vision devices, primarily for industrial use. The field itself is about 30 years old, dating back to work in the 1950s on systems to read printed letters, count blood cells, and recognize military targets.

Despite practical successes, computer vision remains limited, by virtue of inadequate computing power and, more deeply, by the absence of algorithms to visualize adequately a complex and three-dimensional world.

The elements of computer vision are straightforward: A scene is sensed by some sort of imaging device; the images are converted into a digital image, or an array of numbers. The array elements are pixels, or picture elements, with each pixel denoting the brightness of a point on the image scene. A television image is equivalent to about a 500-by-500 array of pixels.

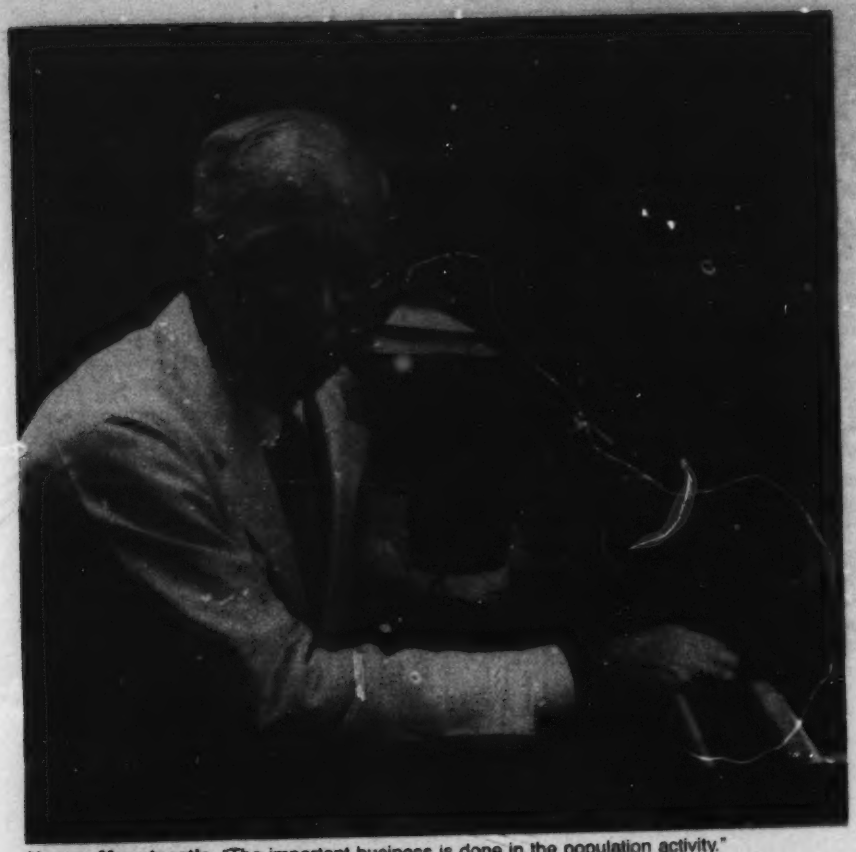
Beyond these basics, the field, and its difficulties, divides according to whether the sought images are two-dimensional or three-dimensional. Template matching is a typical technique for visualizing two-dimensional systems; that is, a printed character—say, the letter *a*—is compared with a representation in the computer's memory. Not all scenes can be anticipated, and the computer is given what amounts to visual elements (such as area to be covered, number of holes), with various schemes (graph structures, for example) used to categorize both features and their likely relationships.

Two-dimensional vision works well, except where the images are poor. Three-dimensional vision systems are, by contrast, profoundly difficult. The image "seen" by the sensors is two-dimensional; sensing and digitizing the brightness of a point say nothing about its distance. Also, at least part of the image is inevitably hidden from and invisible to the sensor.

The various techniques to deal with the first problem rely on shading, texture, and models incorporating shapes, illumination, and the like. For stereovision, a tactic is to identify corresponding points in two images entering the left and right eyes, or left and right cameras, and then, by triangulation, fix the exact location and distance of the point.

Whatever the tactic, the problem is the same, that of "inverse optics." In other words, rather than obtaining the image of a three-dimensional scene, the problem is that of recovering an array of gray-shade values given in the image; or as the late mathematician Mark Kac once put it, the problem is like figuring out the shape of a drum from the frequencies of its sounds.

The second problem—that of the computer completing an image of an object it can see only in part—is being dealt with in several ways. One is to produce images of the object at different orientations and integrate the information. Another is to provide the computer with a model of what it should see and, using the model and what the computer actually sees, have the computer calculate the possible orientations and complete the likely image. ■



Vernon Mountcastle. "The important business is done in the population activity."

pure chance, to reinsert a probe near the equivalent neuron in another brain. A third is to know what is measured. Are signals from a neural network the direct result of the stimulus, coincidental, or the echoed firings of other, undetected neurons, or are they simply noise?

Is there information simply in the fact of two neurons firing synchronously? Further, are network connections constant or transient? Gerstein asserts that "the effective connectivity inside an assembly changes quite drastically with a stimulus change." Poggio argues that "things are much more regular and specific than people think" and that we simply "don't know enough about what individual neurons and synapses do."

The final burden of entry into multineuronal recordings is, quoting Gerstein, that "we're proliferating data at an alarming rate." Further, Mountcastle adds, "down the road we can imagine a monkey having a thousand electrodes implanted. With each of these neurons discharging at 50 times per second, the data flow will be enormous. So, the use of computers in neurobiology is going to increase enormously."

The most immediate information from such recordings of multiple neurons is about what they are doing relative to one another: the behavior of neuron pairs. With 20 neurons, that becomes some 190 cross-correlations—190 basic pictures, Gerstein observes, for that "little chunk of data. You end up with hundreds of little histograms to present to the poor experimenter for interpretation. No one can do that. And no one has yet built an expert system that can do this.

"One of the reasons is that this is not an interesting problem for computer scientists. But it certainly would be interesting for neurobiologists. The difficulty is that neurobiologists recognize only neurobiology as the real science; the tool-making is 'infra dig.'"

Ticking particles

Histograms seem to miss the point of multineuronal recordings, which is to examine the behavior, and meaning, of neural assemblies. The need is for a language, a mathematical language, with which to describe assemblies. Gerstein is exploring one such language, called, for tangential reasons, the gravity method, for describing the state of a neural assembly in computer language. Skipping the intricacies, one can trace the method to William E. Wright, a mathematician who, Gerstein explains, took an informational structure and transformed it into a mechanical one within certain rules. In Gerstein's adaptation, neurons become particles, and their activity time structure, their firing patterns, becomes ticks of attraction.

"Particles that tick together, come together," Gerstein says. That is, by using the gravity method to observe the clustering rate of like-ticking particles, "you can get some idea that, for example, these are pre-synaptic neurons that are feeding post-synaptic neurons in certain ways, because the speed of collapse is related to the strength of connections, and so on. The price of all this is heavy computation. The important point is that this is a way of getting away from pair analysis. Also, we are not bound to long-lasting, static interactions of the neural assembly. What you can really watch is dynamic reorganization. It's not the only idea, and we certainly need other approaches."

Vision

The visible world is created by photons raining upon the retina, a large sheet of some 160 million light-sensitive receptor cells at the rear of the eyeball. These retinal receptors are sensitive to varying intensities of illumination, or gray levels. There are three classes of receptors, each containing a slightly different spectral sensitivity to the various bands of the rainbow. Throughout much of the retina these three types are roughly equally represented, thus forming three spatially distributed systems. What the retina sees, and presumably, what information is passed on to the optical nerve and thence to the visual cortex, is simply a varying continuum of gray levels, of changing light intensities in each of the three receptor systems. It is rather skimpy information, and the problem of vision is how the brain "recovers" a richly painted, three-dimensional world of infinite shapes and textures.

The issue of vision can be teased apart into three strands: What information is available to the brain? What constraints

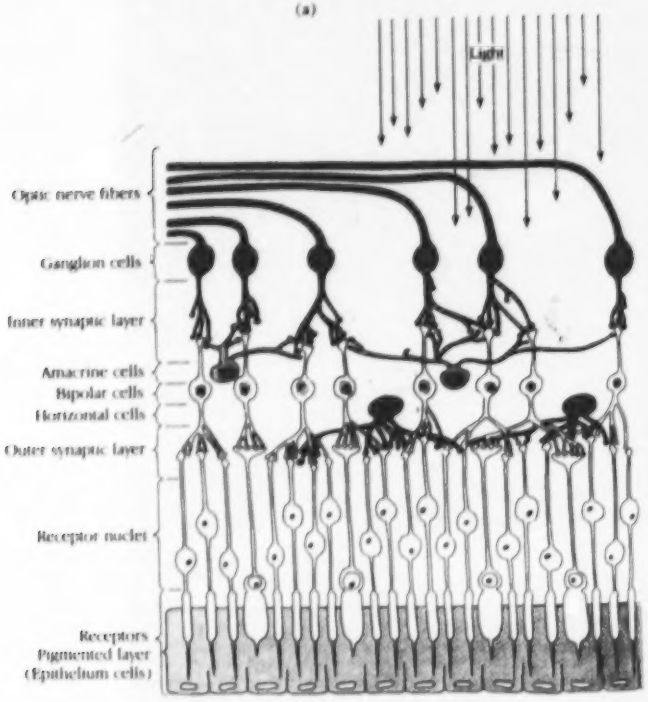
---

Metzger was the author of "A Third Kind of Science," Mosaic *Volume 16 Number 3.*

$$\nabla^2 G * I(x,y)$$

where $\nabla^2 G(r) = -\frac{1}{\pi\sigma^4}\left(1-\frac{r^2}{2\sigma^2}\right)\exp\left(\frac{-r^2}{2\sigma^2}\right)$

(a)



(b)

Vision decoded. A cross section of the retina (below), part of whose function is to compute the mathematical formula (above).

A Marr sampler

"In December 1977, certain events occurred that forced me to write this book a few years earlier than I had planned." That writer was David Marr, then at the Massachusetts Institute of Technology. The book is *Vision*, subtitled *A Computational Investigation into the Human Representation and Processing of Visual Information*. And the "certain events" were leukemia; Marr died in 1980.

Since then, as much as anything can in the conservative world of science, *Vision* has become something of a cult book. That did not happen because the book is easy reading. It is well written, but its territory, that of the way the brain processes visual information, is a difficult one to enter. Its analysis, ranging from the mathematical to the biophysical to the philosophical, is demanding.

It also did not happen because all of Marr's ideas have held up; they have not, although the central ones endure. Rather, the book has gained its audience in part because of Marr's untimely death but more permanently because it bannered quite effectively the usefulness of the information processing prism in crossing the computer/brain barrier. Some random samplings taken from Marr's book *Vision* make the point:

> Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information.
>
> . . . trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers; it just cannot be done. . . . There must exist an additional level of understanding at which the character of the information processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads . . . . It becomes possible, by separating explanations into levels, to make explicit statements about what is being computed and why.
>
> For the subject of vision, there is no single equation or view that explains everything. Each problem has to be addressed from several points of view—as a problem in representing information, as a computation capable of deriving that representation, and as a problem in the architecture of a computer capable of carrying out both things quickly and reliably.
>
> . . . to say the brain is a computer is correct but misleading. It's really a highly specialized information-processing device—or, rather, a whole lot of them. Viewing our brains as information-processing devices is not demeaning and does not negate human values. If anything, it tends to support them and may in the end help us to understand what from an information-processing view human values actually are, why they have selective value, and how they are knitted into the capability for social mores and organization with which our genes have endowed us. •

are imposed? And what does the brain already know? As Tomaso Poggio says, "it is a problem of identifying the correct constraints that you have to use; how to impose them; how to design algorithms that can use them. This is where the real difficulty of vision lies. That *is* the processing problem."

The late David Marr (see "A Marr sampler," accompanying this article) pointed out, "for the subject of vision, there is no single equation or view that explains everything." It was ever so. There are a host of ideas for how the brain computes images of the real world. There is a set of "Shape from ..." or "Structure from ..." processes that attempt to represent as computer images three-dimensional objects from their shading, texture, motion, and the like.

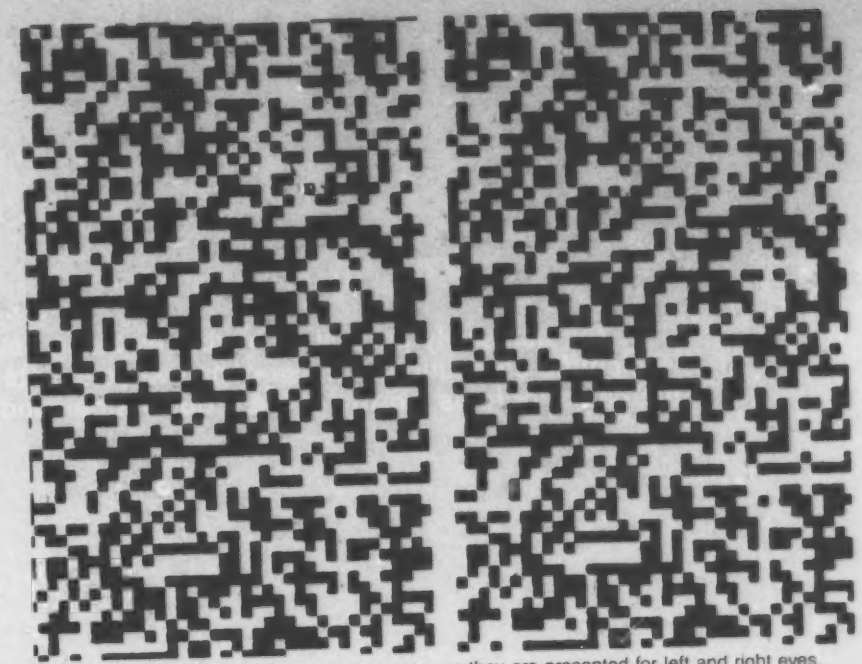
Marr and Poggio have attacked what they called the correspondence problem. Each human eye looks at every point in a view from a slightly different angle, leading to a "binocular disparity." How the brain "knows" that two points imaged by each eye are in fact the same is the correspondence problem, and, as Poggio remarks, "it is the critical one for the computation of stereoscopic and of motion information; the rest is trigonometry"; once two sensors agree that a point is the same, it can be located by triangulation.

That process repeated endlessly creates a stereoimage. The Marr/Poggio approach to the correspondence problem is first to define the invariant information which the brain is given, what Marr called the primal sketch. Second, that raw information is matched among the images with the help of some reasonable, if assumed, constraints. Edges, or sharp changes in intensity value, form the primal sketch. Among the constraints applied to match the edges is that surfaces tend to be continuous and smooth and that points within an edge are unique, so that a point in one image is matchable with only one point in the second image.

A biological complement

Such tactics have had some success, says Poggio, who adds that "the smoothness of physical surfaces can by itself solve the correspondence problem quite effectively. Algorithms have been developed to work on real images."

Granting that machines can solve the correspondence problem, how then can they be made to do it? Neurons are not digital switches, and it becomes difficult to unravel the way the sort of algorithms that endow a computer with stereovision can be solved by neurons. Poggio agrees that "the biological system may make use of several different algorithms, and it may be really difficult to disentangle that."



Binocular disparity. Random-dot stereograms as they are presented for left and right eyes.

However, he notes, the work on how machines can solve the correspondence problem has a biological complement, notably in work at the Johns Hopkins School of Medicine by another Poggio, Gian, who is no relation but who comes from the same town as Tomaso. Gian Poggio works with awake monkeys and random-dot stereograms.

Random-dot stereograms were invented by Bela Julesz of AT&T Bell Laboratories. Julesz created them by filling a sheet of paper at random with black and white dots, making a copy, and then shifting to the left and right square areas of dots in each piece of paper. The spaces left by the shifts were filled by more dots.

Under normal conditions, all that is apparent are two sheets of paper with random doodles. However, when a stereoscope is used so that each eye looks at only one piece of paper, a perfect square appears to float above the stereogram. Julesz' work showed that (1) the brain did not have to recognize an object first in order to see it three-dimensionally and (2) the brain can compute a stereoimage even when one isn't readily apparent; it has the ability to solve the correspondence problem quite easily.

Gian Poggio has recorded the response of neurons of the visual cortex of monkeys looking at such stereograms. He has found that specific neurons will fire when they can match two points at the same depth in the stereogram; further, different neurons will respond to different stereogram depths. As Gian Poggio notes, "these neurons perform the correspondence solution."

The biological and computational work is now in a synergistic mode, each part prompting the other, each pointing to similarities and differences in how a brain and a computer solve the correspondence problem. For example, Gian Poggio finds that neurons have little difficulty in dealing with large disparities in stereogram depths, while "some vision algorithms tend to fall apart."

The root question of whether the algorithms running in the computer or the brain will turn out to be at all alike is really immaterial. The point, as Tomaso Poggio says, is that "if you want to understand the brain, you have to understand it also from the point of view of information processing. This level of understanding is different from and complementary to understanding how individual neurons behave and how they are connected." •

by Arthur Fisher

The massively parallel architectures being devised for scientific computing promise cognitive scientists for the first time a set of tools worthy of the problems demanding attention.

The heart has its reasons, which reason knoweth not.

Blaise Pascal, *Pensees*

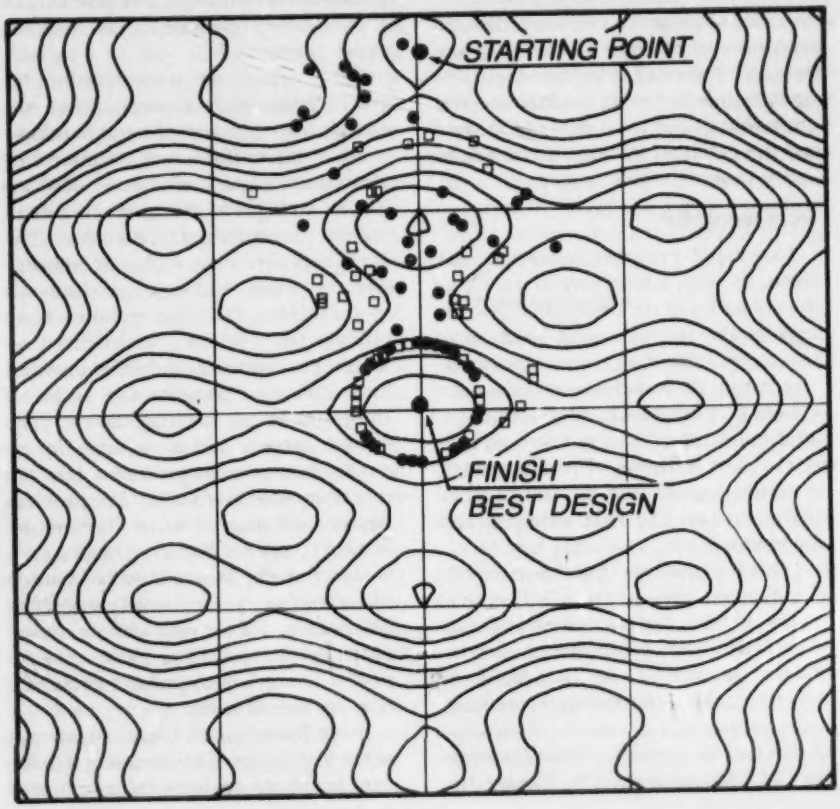
At Stanford University, cognitive scientist Roger N. Shepard uses a computer to generate perspective drawings of three-dimensional objects in space. Then, he measures the speed with which people can rotate objects mentally and observes that the time required is proportional to the angle of rotation. The goal: to adduce a model of spatial imagery in humans.

At the University of Wisconsin Medical School, Robert B. Friedman is designing an interactive, computer-assisted instruction module for an eight-hour course on the basic science and clinical aspects of immunology, developed to incorporate the fresh perspectives of cognitive psychologists on medical education. The course will be tested against the conventional teaching approach: structured passive learning.

At Rutgers University, biophysicist Evangelia Tzanakou confronts an anesthetized frog, with shifting patterns of dark and light squares. A hair-thin electrode is implanted in one of the frog's brain cells. The animal's changing responses to changing patterns are evidenced by altered voltages generated in its brain. The shifting patterns are computer-generated. A special algorithm enables the computer to sift through what has gone before and generate new patterns, not at random, but in a sequence that will increasingly maximize the frog's response. The object: to help unravel the still mysterious mechanism of visual pattern recognition in amphibians—and eventually in humans.

At the Georgetown University Medical Center, an instrument called the *Cenog* helps diagnose neurological disorders including such learning disabilities as dyslexia and other reading problems. The name stands for "computerized electro-neuro-ophthalmograph." The system tests responses to light or other signals directed at patients seated in a special chamber. The computer automatically varies the stimuli according to the subject's eye movements.

At The Carnegie-Mellon University, Geoffrey Hinton, Scott Fahlman, David Touretzky, and their colleagues are investigating a computer model called the *Boltzmann Machine*. The model has a massively parallel architecture and is designed, in part, to operate as a network of neurons. The model performs an equivalent of learning and solves recognition and other problems characteristic of human intelligence. The dual



The best design. Starting from an imperfect design (left), Ihor Bohachevsky's program follows the path (right) of improved design—solid circles—from starting point at top center to best design at center.

objective: to build more intelligent computers and to better understand the processes of human cognition.

These experiments represent a sampling of the uses to which computers are being put in the service of cognitive science today. Cognitive science itself is a relatively new discipline, surfacing as an important entity in the late 1970s. Or rather, it is interdisciplinary, since it draws on and, in a sense, links computer science with psychology, linguistics, philosophy, neuroscience, and anthropology. Precise definitions of its purview, and avoidance of overlaps with its parent disciplines, are obviously difficult. But its affinity for computers and computing really is inherently logical.

According to a National Academy of Science study in 1983 titled *Report of the Research Briefing Panel on Cognitive Science and Artificial Intelligence*, "Cognitive science and artificial intelligence stand together in taking information processing as the central activity involved in intelligent behavior and in taking the framework of modern computer science as the foundation for understanding information processing systems. Cognitive science reflects a concern with how humans process information . . . Artificial intelligence reflects a concern with how computers process information. . . . The rapid growth of these disciplines in recent years is advancing our understanding of the nature of the mind and the nature of intelligence on a scale that is proving revolutionary."

Terra Incognita

This rapid progress, however, still leaves us with a long way to go. Of all the mysteries in the universe, the workings of the human mind seem most veiled, most baffling, most intractable. The nature of consciousness, thought, learning, perception, and language, after millennia of introspection and the minings of millions of philosophers, mentalists, psychologists, and psychiatrists, remains to a great extent a terra incognita.

What is a conscious mind anyway? As the Oxford philosopher J. R. Lucas wrote in 1961, "At one's first and simplest attempts to philosophize, one becomes entangled in questions of whether when one knows something, one knows that one knows it, and what, when one is thinking about oneself, is being thought about, and what is doing the thinking. After one has been puzzled and bruised by this problem for a long time, one learns not to press these questions."

Touretzky. With Hinton and Fahlman: A network of neurons from a *Boltzmann Machine*.

Other gaps in our understanding become distressingly apparent when scientists and engineers attempt to endow nonhuman artifacts with human-like mental attributes, whether called artificial intelligence or something else.

Take perception, for example. Psychologists now know that an infant only days or weeks old can recognize its mother's face and distinguish it from others. Yet years of sophisticated research and millions of dollars invested have failed to transfer the sense of sight—the ability to recognize everyday ordinary objects and distinguish, for example, between a tiger and a lion—to robots or other machines. "The problem has proved much harder than we assumed," says Dr. Tomaso Poggio, a researcher at the Massachusetts Institute of Technology's Artificial Intelligence Laboratory. "It's obvious why we underestimated the problem; it's so easy for people to see." But we don't know just how we see so easily.

David Rumelhart, a cognitive scientist at the University of California at San Diego, is concerned with the question of how educated people read words so easily. "How is it," he says, "that people can rapidly identify the meaning of a visual symbol?" Even more provocative is "Why are people with large vocabularies quicker at recognizing simple words? What sort of information retrieval system gets better as its data base increases?"

Similarly, anyone, even a very small child, can isolate a particular set of compressions and rarefactions of air, what we call sounds, from a welter of others, to zero in on one particular voice and to then interpret those sounds as language. The phenomenon, the isolation even of one's own name from such a welter, has been called the cocktail party effect. Yet giving a computer the ability to understand a single person, speaking slowly and distinctly, has been a long and difficult struggle.

The nature of human language abilities has been one of the knottiest of all problems in cognition and is at the core of the effort to create artificial intelligence. It is like a carbuncle on the neck of AI specialists, who cannot perform in their constructs what any normal post-toddler can in the way of language. It is this facility, of course, that seems to separate humans not only from machines but from all other living things.

"There may be capabilities which the human organism is uniquely prepared to

---

Fisher was the author of "New Languages for Old," Mosaic Volume 16 Number 3.

develop," says Gregory A. Kimble, a psychologist at Duke University. Language is the clearest example. Although investigators have had impressive successes in teaching chimpanzees to perform language-like behavior, even more impressive is the difficulty of such training.

"The human two-year-old picks up language effortlessly and without any special program of training," says Kimble. "For the chimpanzee it takes extensive tutelage. It is as if the human being is prepared to develop linguistic skill in a way that the chimpanzee is not."

The ability to deal with nonlanguage problems also seems uncanny. People attempting to move chairs through narrow doorways, for example, have usually preimaged the maneuver so that they automatically know which way to orient the legs. How does the brain perform this piece of mental gymnastics? Is spatial imagination governed by laws that are as strict and logical as those of Euclid's geometry?

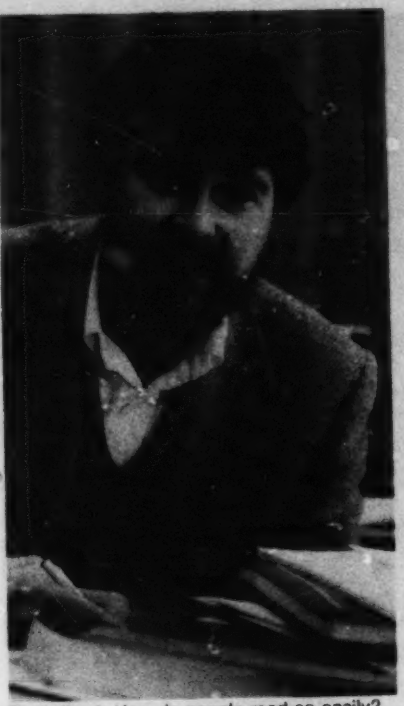
Simulation and analysis

To solve such problems, which, says Rumelhart, are at least as complex as those of the natural sciences, cognitive scientists are making intensive use of the computer in various ways. With its superhuman speed and number-crunching ability, the computer serves nobly as an aid for controlling, collecting, and analyzing reams of otherwise unhandleable experimental data.

But there are more creative uses as well, and more interesting. One is to simulate human mental processes with computer models and then to use the computer to test the model's adequacy. Work in this area is being carried on at a ferocious rate, not only better to understand human cognition but to produce a breed of computers with human-like cognitive abilities. The validity of the so-called brain/computer metaphor, on which much of this current line of research rests, is being hotly debated.

Much less controversial is a second major application of computers: generating carefully tailored stimuli and analyzing a subject's response.

The mind of a child, said John Locke (1632-1704), is a tabula rasa upon which experience writes to create the process of learning. Experience comes in the form of stimuli from the world outside. And for many years, students of cognition have observed people's reactions to given stimuli as a means of probing the



Rumelhart. How do people read so easily?

workings of the mind. One of the founders of modern psychology, Edward Chase Tolman (1886-1959), wrote in a classic description of the process in 1922, "The stimulating agency may be defined in any standardized terms, those of physics, of physiology, or of common sense, and it constitutes the independent, initiating cause of the whole behavior phenomena. . . .

"The behavior act is simply the name given the final bits of behavior as such. The behavior act together with the stimulating agency constitutes the fundamentals upon which the rest of the system is based. . . . They alone . . . tell us all that we know of . . . an organism's mentality (even when that organism is another human being who is capable of introspection)."

In cognitive science, says Paul Rosenbloom, psychologist/computer scientist at Stanford University, computers are extensively used for running stimulus-presentation experiments. The researcher plants the subject in front of a terminal; the computer presents a stimulus, records the response, and then runs everything through some kind of data-processing program.

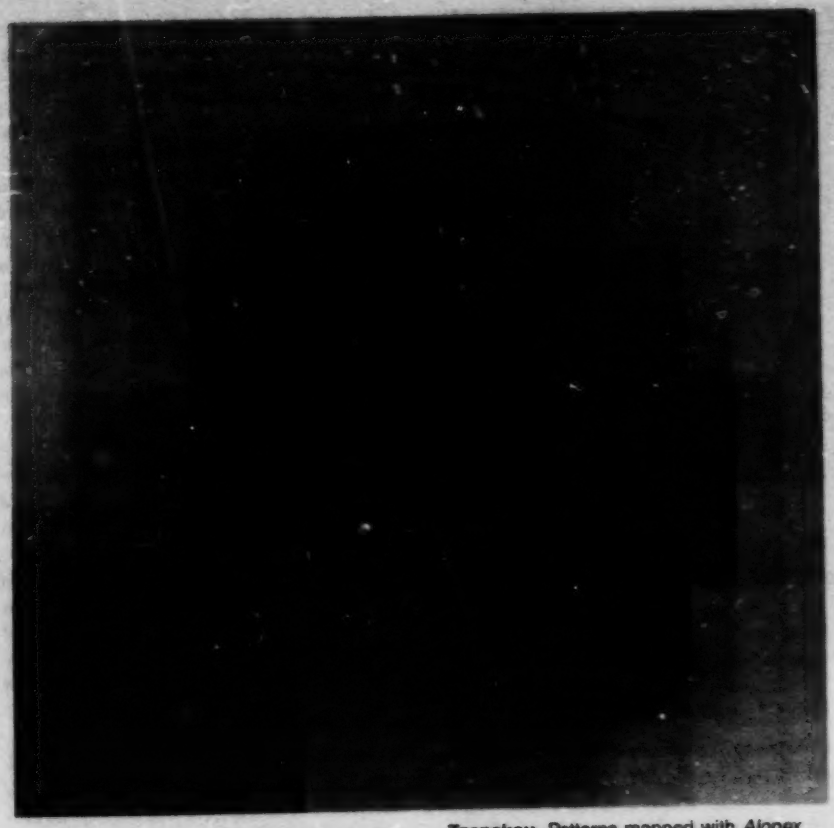
One typical example is the type of experiment called lexical decision tasks that investigates the way people understand words. A string of characters is presented on the screen and the subject is asked to press one of two buttons, depending on whether the subject perceives the string as a word or not. "We can use these experiments to help understand various underlying factors," says Rosenbloom. "For example, you can present a related word just before the word presented to the subject, or just before a nonword. How long does it take in each case? If you present the word *bat* and then the word *ball*, you might expect the subject to recognize that *ball* is a word faster than he or she would if you had first presented the word *restaurant*."

You could also, says Rosenbloom, try sequences that would test whether people are aware of word ambiguities, like the two meanings of the word *bat*. Such ambiguities are one of the major obstacles to endowing intelligent computers with natural language abilities. "These experiments help us understand how words relate to one another and how they are represented in the mind. And the computer plays an essential role."

A different kind of stimulus presentation, and a different kind of subject, is used in the perception research being carried out at Rutgers University. Beginning with the pioneering work of Jerry Lettvin, now at MIT, and his colleagues in 1959, experiments in visual perception have turned to the frog's visual system because of its simplicity and accessibility. In 1959, Lettvin and his associates moved various objects in front of a frog's eyes while an electrode implanted in one of the fibers of the frog's optic nerve recorded changes in electrical response.

Pattern extraction

Today a version of this experiment is being performed, but with sophisticated computer science as an aid. In Evangelia Tzanakou's laboratory at Rutgers University's College of Engineering, an individual brain cell in an anesthetized frog's visual center is monitored by means of an electrode, as a computer-generated pattern is shown to the frog. To learn something useful, the experimenter must weed out unproductive patterns from good ones. Good patterns will stimulate a powerful response as measured by the electrode. But the number of possible variations of light and dark patterns is virtually limitless. It would take a lifetime or more of research to test them randomly.



Tzanakou. Patterns mapped with *Alopex* measure potentials from a frog neuron.

To solve this problem, Tzanakou developed a computer algorithm called *Alopex*, for algorithmic logic of pattern extraction. She presents one pattern, such as a checkerboard pattern with a random distribution of light intensities, of white, gray, darker gray, and so on. "That's the image the frog sees," she says, "the signal from the outside world. When this information reaches the neuron we are monitoring, it is already partially processed by the computer. What we get through the electrode is an action potential we can measure. The total number of these potentials elicited by each pattern we then send to the computer."

The computer generates random patterns, which are, however, closely related to the first one. Using the *Alopex* algorithm, it compares the potentials elicited from each pattern, at the same time making a detailed, point-by-point comparison of the patterns themselves. The computer continues to generate patterns in an increasingly predictive manner, so as to maximize the frog's response and single out the most evocative patterns.

"In effect," says Tzanakou, "the neuron tells you what it likes. As you go on, the pattern changes aren't random any more. If the response goes up when the light intensity goes up a little in one spot, the computer says 'Okay, let's increase it a little bit more and see what happens.' If the response increases continuously, you know you are going in the right direction. But if the response goes down beyond a certain point, you know you have exceeded the optimum intensity for that spot in the pattern. The computer changes the whole pattern for you automatically."

Another level of research involves experiments to develop and test a variety of models of human cognition. Roger N. Shepard and Lynn A. Cooper have been collaborating intensively on the problem of how people manipulate images of objects in their minds, following a line of inquiry begun by Shepard in 1971. Shepard is a research psychologist at Stanford University; Cooper is a psychologist at the University of Pittsburgh and senior scientist at its Learning Research and Development Center. They believe that the mechanics of spatial imagery has well-defined, statable rules, and the experiments they carry out tend to fit well with their model.

Angular rotation

In Shepard's and Cooper's experiments, subjects are shown pairs of computer-generated perspective line drawings. Each drawing shows a three-dimensional solid object consisting of ten cubes joined at their faces to build an arm-like construct with three elbows, or right-angle bends. Some of the pairs were of spatially identical objects, but with each member of the pair shown in a different orientation. Other pairs were identical except that they were enantiomorphic, or mirror-images. Again, each member of the pair was shown in a different orientation.

Because the objects in each pair gave no obvious clues, it would appear that the only way subjects of the experiment could decide whether the objects in each pair were identical was to mentally rotate one until it had the same orientations as its partner, then mentally see if they matched. Could this specific model of spatial imagery—turning things over in the mind, as the investigators call it—be verified by a set of objective, quantitative results?

The answer is yes, by means of a clever experimental stratagem. The angular rotation needed to bring each member of a pair into alignment was not the same for each pair of objects. Some required very little rotation, some the maximum rotation of 180 degrees, others values in-between. The time for each subject to reach a match-or-no-match

decision, counting from the instant the drawings were displayed, was carefully clocked. If the model of Shepard and Cooper was correct, then the greater the degree of rotation, the longer that it should take to turn things over and reach a decision.

Result? The reaction times, as predicted, grew linearly with the angular difference between the orientations displayed. Subjects needed only a second to confirm that like objects were indeed identical when their orientations were the same. As the angle of rotation increased, so did the response time, until a maximum of an average 4.4 seconds was reached for 180 degrees.

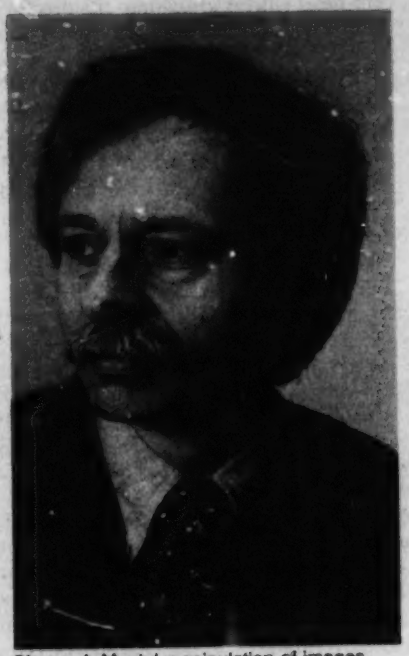
For years, a great many researchers in human cognition have sought to use the computer not merely as a tool for generating stimuli, collecting and analyzing data, and testing models of information processing, but as a paradigm of the human brain itself. And from the opposite pole, as it were, a host of researchers in AI have sought to create more sapient computers and computer-run machines by modeling them on what we know of the human brain's architecture and functioning.

Ferment and debate

These efforts have generated a decade-long, ongoing debate that has contributed a rich intellectual ferment to both sciences—cognitive and computer. Is it valid to set up a parallelism between the human brain and a nonhuman computer, no matter how potent? Should either cognitive or computer scientists attempt to delve into their own mysteries by borrowing from the data base of the other camp? In essence, is the brain/computer metaphor valid?

At least two doyens of the AI community take a relatively positive view. Roger Schank, who heads Yale University's Artificial Intelligence Laboratory, says, "We believe that AI is impossible unless we understand how people think." And Patrick Winston, director of the Artificial Intelligence Laboratory at MIT, has said flatly, "There's no reason a computer can't exhibit as much intelligence as a human. Neurons or silicon, the difference is irrelevant."

But other computer scientists are more skeptical. Says David Eliot Shaw of Columbia University's department of computer science, "There's still a great deal we don't know about human thought. We're a long way from building a machine capable of competing with the human brain."



Shepard. Mental manipulation of images.

Other scientists resist relying on the brain/computer metaphor at all; indeed they spurn any attempt to build desirable properties into a computer by modeling the human brain.

Ihor Bohachevsky is a mathematician with the Advanced Weapons Technology Office of the Los Alamos National Laboratory. He and his colleagues are developing a code that will think its way through a multitude of possible lens designs and come up with a near-perfect lens every time, one that will fulfill exactly certain requirements. To be successful, Bohachevsky says, this AI application must be able to learn from experience, just as humans do, blending new information into its memory for future use. The application will avoid expensive, time-consuming experimental verification, and Bohachevsky believes that within five years it could be used to design a host of near-perfect products, ranging from tires to microchips.

The team believes that if they store enough information about the object they want to design, in this case a lens, the computer will be able to progress from one bit of data to the next, finding the best move at each stage and continually improving the design.

"We're trying to develop an intelligent approach," says Bohachevsky. "We don't just look up a million things and compare them. We use what the computer already knows and let it find the answer."

But Bohachevsky's group is not trying to imitate the human brain to derive this intelligent approach. "The brain has not evolved for the purpose of solving technical problems," he says. "If you wish to build a flying machine, then the worst possible thing you can do is to imitate birds. That delayed flight for a least a thousand years. That's why I think trying to imitate the human brain is inappropriate. You don't have to understand how the brain works, only how a computer works," Bohachevsky declares.

Agnosticism

It is also possible to find early supporters of the brain/computer metaphor who have become, over the years, more skeptical, or at least agnostic. Linguist/computer-scientist Terry Winograd now believes that between what we know about the way the mind works and what may actually be going on inside the brain there is a gulf so wide that we may never bridge it.

Winograd serves as consultant at the Xerox Corporation's Palo Alto Research Center and does research in artificial intelligence, computational linguistics, and cognitive modeling at Stanford's Center for the Study of Language and Information.

One striking example of the gap in our knowledge of the mind's workings, and the consequent inability to impart certain mental attributes to computers, is the question of understanding language, an area Winograd has probed deeply. Natural languages, the ones people speak and write, have built-in ambiguities of several types. As a result, the meaning of a sentence in a natural language often depends not only on the form of the sentence but on its context, which is needed to resolve the ambiguities. Humans solve these problems without apparent effort, by a process whose mechanics remain incompletely understood, which is to say, magical. No computer program to date has been able to deal with them with complete success. In fact, they constitute formidable obstacles to many kinds of intelligent computer applications, including speech recognition and translation from one natural language to another.

Winograd has set up a kind of hierarchy of language ambiguities and has provided examples of each. Any literate

person can easily supply his own favorite examples. Winograd's categories extend from lexical ambiguities to structural, deep-structural, semantic, and programmatic ones.

Ambiguities

In lexical ambiguity, the simplest kind, a single word such as *bat* can have two or more meanings. If the sentence reads "The second baseman picked up his bat," the meaning to a human reader is unambiguous. But the sentence "Stay away from the bank" might be advice to a child playing near the river or to an investor about to make a deposit. Only the context—what is outside of but embeds the sentence—can reveal which sense is correct.

In structural ambiguity, there is a choice between two grammatical structures. For example, "He saw that gasoline can explode." Did he understand a possibility or witness an explosion?

Deep structural ambiguity is even more subtle: Two readings of the same sentence can have the same ostensible grammatical structure but totally different meanings. Take the sentence "The chickens are ready to eat." Obviously, something is ready to eat something else. But are the chickens or their chicken feed about to be eaten?

Speech recognition

The impulse to communicate by speech with a kind of robot secretary found expression in Karel Capek's seminal 1921 play, *R.U.R.*, in the person, if one can call it that, of Sulla, a machine that takes oral dictation and types a transcript. The nonfictitious search for a mechanism capable of speech recognition has been going on for decades and has proved one of the most frustrating of all efforts to reproduce a quite ordinary human cognitive faculty. In effect, the ability to comprehend spoken language may be even more difficult to analyze and replicate via computer than the mechanisms of human visual perception.

Yet the race is on, says Michael L. Dertouzos, director of MIT's Laboratory for Computer Science. "Whoever breaks the logjam to make intelligent multiprocessor machines that can, for example, understand spoken words will gain control of the information revolution." He observes that machine recognition of speech is a principal goal of Japan's fifth-generation computer program. And the U.S. Department of Defense's Advanced Research Projects Agency has been sponsoring work on advanced speech-recognition systems off and on since 1971.

The potential rewards of such an achievement are obvious. Data and instructions could flow to a computer far more rapidly than at the rates now accessible to most users. The 150-words-per-minute that is a typical conversational flow is about double the rate of most good typists. For general business alone, that is reason enough to develop a so-called talkwriter.

Beyond that are many situations in which voice input is desirable, even essential. Airplane pilots in emergencies. Tank commanders under fire. Handicapped persons able to speak but unable to manipulate objects in their environment. An even greater benefit might be a deeper understanding of the still-obscure pathways of spoken language comprehension itself.

Obstacles to unfettered conversation between man and computer are formidable. True, there have been limited successes in the evolution of speech recognition systems, ranging from the trivial, such as arcade video games that respond to specific spoken cues, to the practical, such as speech controls for auto radios, voice inputs for assembly-line pickers in factories, and telephone banking services that can respond to simple vocal commands. But most of these applications have tiny vocabularies, are limited in combinatorial power, require exaggeratedly long pauses between each word, and generally have to be dedicated to a single individual and thus be incapable of recognizing any other person's voice. The goal of talking to a computer the way you would talk to a person, says Stephen Levinson of Bell Laboratory's acoustic research section, "is a long way off."

By the standards of human intercourse, what researchers in the field would like to accomplish is not unusual. A computer should be able to recognize, with high accuracy, human speech that is continuous rather than separated by unnatural pauses, contains a large vocabulary, and is independent of the regional accents, peculiarities of phrasing, pitch, and other such quirks of the user. It would also be nice not to have to isolate the speaker in an anechoic chamber or even a very quiet environment. After all, a human listener successfully interprets the utterances of thousands of different individuals, with accents ranging from Deep South to Bronx, inflections from Brahmin to gutter, pitch ranges from basso profundo to squeaky, tone-dependencies from Japanese to Chinese, and clarity of enunciation from nightly news anchorman to tobacco auctioneer. Furthermore, a human listener can pick out the distinctive sound and sense of an acquaintance's voice from the din of a boiler factory or the babble of a cocktail party.

Constraints

Why should attaining anything remotely resembling these human abilities in a computer be so difficult?

If computer recognition of speech were simply a matter of matching the acoustic signature of a word, or of the discrete sound segments called phonemes that constitute a word, against a sort of electronic template stored in memory, the task would not be so hard. In fact, that is exactly the procedure used in the primitive systems that already represent what would laughingly, if computers could laugh, be called contemporary successes. Such systems are constrained by very small vocabularies, perhaps 100 words or less, separated by pauses and trained to the idiosyncracies of specific speakers. Their success is based on relatively straightforward signal processing.

Real speech is not so simple. Take vocabulary, for example. There are more than 400,000 words in the English language. A reasonably educated adult may use 10,000 to 15,000 words in casual conversation but possesses a reservoir of 100,000 words, and such a person may freshen his conversation with any one of these on occasion. Based on template matching alone, enormous computing power would be required to cope with this number.

Semantic ambiguity arises when a phrase can play different roles in the overall meaning of a sentence. "David wants to marry a Norwegian," for example. Does David want to marry a particular woman named Ingrid, who happens to be a Norwegian? Or is he casting about for any old Norwegian? The difference, according to Winograd, is whether the phrase "a Norwegian" is referential (to Ingrid) or attributive.

Pragmatic ambiguity results from the use of pronouns and such special nouns as *one* and *another*. In the sentence "When a bright moon ends a dark day, a brighter one will follow," what will follow, a brighter day or a brighter moon? Easier for a human to resolve is the ambiguity in "She dropped the plate on the table and broke it."

The point Winograd makes is that many ambiguous sentences do not present problems to human readers or listeners because they bring to the task an understanding of the context. In "The porridge is ready to eat," a human knows that porridges are abstemious; the computer does not. The difference leads machine translation systems into such notorious bloopers as the recurrent term *water goat* in an engineering treatise on hydraulic rams.

Will such difficulties make the brain/

Nor is that the worst case. Because different speakers produce entirely different acoustic signals for the very same word—compare *Harvard* as spoken by a Bostonian and by a Texan—elementary template-matching systems are speaker-dependent. The computer is actually taught to recognize, word by word, the speech of a particular individual. Should the speaker develop sinusitis and a hacking cough, the computer might very well fail to make a match. Obviously, the larger the vocabulary, the more tedious and time-consuming this training process becomes. To render a speech-recognition system speaker-independent thus becomes an entirely different kind of challenge.

Doing away with the third constraint—achieving the recognition of continuous speech—poses even thornier problems. The acoustic properties of words and phonemes in connected discourse are not only different from those of isolated words, they differ according to variations in the words that surround them, the phonetic environment. For example, the basic phoneme *t* will yield very different acoustic signals depending on whether it is spoken in *tree, stern, beetroot, tie,* or *city*. This confusing effect is called coarticulation; it can occur both within words and at their boundaries, as in *gas station*. The word *this* will be different acoustically depending on whether it is used in *this car* or *this ship*.

The human hearing apparatus and brain manage to cope with the perplexities of run-on speech and with other perplexities as well. So far, no computer system has been able to come close. The reason is that a true speech-recognition system, if it is to operate without constraints, must know the speaker's language.

Toward a science

According to Victor Zue, an MIT scientist at the forefront of speech-recognition research, the field is in its infancy. "The development of advanced speech-recognition systems will require expertise in signal processing, system theory, and artificial intelligence, built on a solid understanding of speech science. Most of the current systems utilize little or no speech-specific knowledge; rather they derive their power from general-purpose pattern-recognition techniques."

Work on advancing beyond mere pattern matching is now proceeding at many laboratories. At Kurzweil Applied Intelligence, in Waltham, Massachusetts, Raymond Kurzweil is developing a *Voicewriter* that would be a speech-to-text system with a 10,000-word vocabulary. It would be speaker-dependent and noncontinuous but would be able to take 150 words of dictation per minute. It will rely heavily on linguistic and speech algorithms and parallel expert systems. One system might, for example, begin the process by recognizing phonemes, another would parse sentences, another predict what parts of speech might reasonably follow another. Present plans call for seven such experts to operate simultaneously in a complex arrangement of parallel processors.

Another and totally different tack has been under pursuit for some twelve years by a team led by Frederick Jelinek at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York. The IBM system spurns template matching and relies instead on a two-stage statistical modeling procedure. The first model samples speech sounds 100 times a second. Based on a 20-minute standardized sample the speaker has previously read to the computer, the system calculates probabilistically what word the speaker is likely to have uttered. A second model, produced through analysis of some 25 million words of business correspondence, predicts what English word is likely to follow the two previous words.

The IBM system has demonstrated an accuracy of 95 percent in comprehending a 5,000-word vocabulary, an impressive achievement. Jelinek likes to display the system's ability to deal with the complexities of natural English language by reading it Hamlet's line "To be, or not to be." The system must decide among *to, too,* or *two*, and also among *B, be,* or *bee*. It does so successfully. However, even at its present stage of development, the system uses an IBM mainframe computer (the 4341), three Floating Point System Array Processors, an IBM PC, and other special hardware. But Jelinek thinks that in five years a compact system could be feasible.

It will be many years, though, before all the goals of human/computer conversation will be attained. Artificial-intelligence expert Edward Feigenbaum conjures, as an example of tiger traps that lie ahead, a dialogue between a general in a war-room and the supremely intelligent computer that, by the year 2020, is supposed infallibly to manage Armageddon for him.

Computer: General, enemy missiles are approaching!
General: By land or by sea?
Computer: Yes.
General (furiously): Yes, what?
Computer: Yes, SIR! •

A.F.

computer metaphor insupportable? Stanford University's Paul Rosenbloom, who was rapporteur for the National Academy of Sciences report on cognition and AI, sums up the controversy this way: Today, he says, "the brain-computer controversy is hot, with strong beliefs on both sides. On the AI side, you find a variety of philosophies, but the dominant one is this: Why should there be anything similar in the way people think and the way you'd want a computer to do things? It's not even worth looking at people; it might be misleading. So you just sit down and write the best computer programs you can for the job.

"Similarly, on the cognitive science side you find people who say the way the mind works is not at all the way a computer works, so it's totally ridiculous to think of the computer as a model for the brain."

Rosenbloom ranks himself with those who take a middle position. "We have only one existing instance of an intelligent organism that we can understand at all," he says, "and that's people. And we're much better off—it's a good heuristic—if what you do is based on that example, as a guide to what you are doing right.

"In cognitive science," says Rosenbloom, "the middle position holds that a computer as normally programmed is certainly not a good model for the brain. But you can build models based on concepts of information processing, as we do in AI, build them within the computer, and make what I think are quite accurate models of things that are going on in the mind."

A pool effort

One such modeling attempt is the program to build the so-called *Boltzmann Machine* at Carnegie-Mellon University. The project pools the efforts of computer scientists Geoffrey Hinton, Scott Fahlman, and David Touretzky, among others. It exists for now only as software simulation of a theoretical computer, one with an extraordinary architecture. It borrows from the notion that the brain has billions of processors called neurons, is highly parallel, and has a plethora of interconnections. It is the product of much thinking about the nature of the brain, "a poor thing," Touretzky comments, "but our own.

"After all," says Touretzky, "the brain is working with severely limited hardware. Neurons are very slow; they fire less than once per millisecond. And they're very unreliable. They get tired, they die. Obviously, the brain must make use of some very clever organizational principles to take advantage of its massive parallelism."

Winograd. A wide gulf between what is knowledge about the mind and actual brain activity.

In fact, one advantage of the brain's organization is that information is distributed over many cells, so that even though thousands of brain cells die every day, memory is not lost in an otherwise healthy brain.

The concept of the *Boltzmann Machine* is to have many processors, roughly corresponding to the brain's neurons, but far fewer in number, with each connected to hundreds of others. According to Touretzky, the behavior of this neural network can be analyzed using statistical mechanics, in particular the Boltzmann distribution, which describes the behavior of a gas at thermal equilibrium.

The distribution is named for Ludwig Boltzmann, who lived from 1844 to 1906, an Austrian physicist whose work on the thermodynamics of gases provided the foundation for modern statistical mechanics. Information is stored not in the processors but in their connections, which are assigned different numerical weights that play a role in statistical decision making. Although there are rough analogies to neurons and synapses in the Boltzmann architecture, the model does not represent a true biological picture of the brain. For example, the processors are either on or off, merely add inputs, and have a fairly limited number of possible connections. These characteristics are not shared by real neurons, which among other things can handle more than 50,000 synaptic inputs each and engage in complex information processing.

Nevertheless, simulations of the *Boltzmann Machine* have demonstrated that the model is capable of some relatively advanced cognitive tasks, including learning, pattern matching, and programmable symbol processing. Thus the simulations have confounded some critics of Boltzmann-based theory who have held that such systems would be incapable of general-purpose symbolic reasoning. "The brain's success," says Touretzky, "depends on organizational principles that are as yet unknown. By exploring the problems of computing with distributed representations, computer scientists may eventually uncover some of these principles."

The computer gap

Progress in this area, as in many other fields of cognitive science, faces a growing problem, one that could succinctly be called the computer gap. According to David Rumelhart, the whole class of problems, including models of perception and cognition based on the neural architecture of the brain, could benefit from access to high-speed vector-based machines because they use algorithms that are computationally intensive, can be described as semilinear, and resemble those used in problems in fluid dynamics and other physical systems.

The high-speed machines Rumelhart yearns for are the so-called Class VI supercomputers such as the *Cray* XMP and the *Cyber 205*. Their combination of processor speed and memory size gives

them 20 to 50 times the computing power of current mainframes such as the IBM 3033. Moreover, these computers include pipeline arithmetic units capable of extremely efficient computation of vector arithmetic.

Writing about the constraints on cognitive research imposed by the lack of access to adequate computing facilities, in a conference report on advanced computing for psychology issued in May 1985 by the Federation of Behavioral, Psychological and Cognitive Sciences, Rumelhart states that the models that can be investigated on available computers are too small in scale to be realistic. As a result, existing models focus on a single aspect of a much larger cognitive system.

Recent advances, Rumelhart writes, have shown that introducing randomness into cognitive models can actually improve their performance. But using Monte Carlo simulations, a standard technique for introducing randomness, can require staggering numbers of simulations. "Current simple models now run slowly on *Vax's*, about 1,000 times as slowly as the real time processes they attempt to simulate," says Rumelhart. Using simulation techniques or other critically needed extensions of the modeling process would slow the acquisition of results "by at least another two orders of magnitude."

It is clear that an increase of two or three orders of magnitude in speed, which seem possible with Class VI supercomputers, would be enormously helpful in tackling all these problems.

Dertouzos. "Whoever breaks the logjam . . ."

Rosenbloom. "Brain-computer controversy is hot, with strong beliefs on both sides."

Intolerable constraint

A shortfall in computer speed also makes promising lines of research utterly impossible. Margaret Hagen of Boston University uses computer graphics as a vital research tool in her work on visual perception. "With today's improved state of the art in graphics," she says, "vision scientists have the opportunity, never available before, to advance their science."

Computer-generated stimulus displays, she explains, have three characteristics that determine the feasibility of creating an adequate display with the hardware and software available to the moderately well-funded researcher. The three are resolution, scene complexity, and degree of freedom from artifacts. Certain applications, she says, can be carried out successfully with available resources "as long as the stimulus complexity is minimal, the number of images is few, and the presence of artifacts and mediocre photorealism is tolerable."

Unfortunately, in certain areas at the forefront of modern psychophysics, Hagen says, these constraints are simply intolerable: Images must be realistic and/or produced in great number, with a resulting vast increase in the computer time needed to generate the stimuli. "A high-resolution film (4000 by 4000 pixels) requires ten hours of computation per frame on our IBM 3081D, or 8.2 years for a ten-minute film. "No researcher can wait eight years to create a stimulus display for one experiment."

Applications hampered by this kind of computer gap, Hagen says, include investigations of motion perception and of spectral and achromatic surface reflectance.

Another large class of problems in cognitive science requires the manipulation of symbols rather than arithmetic operations and may be even more tantalizingly difficult, close to the nub of the human mind, more profoundly demanding of intelligence. Rumelhart poses as examples of this kind of problem modeling language (the prototype example of a symbolic system), attention, decision, storage and retrieval of information in memory, and problem-solving. These areas are not well suited to the architectures of most current computers, including the Class VI supercomputers, whose vector processing is especially suited to massive number-crunching. "Some current vector machines," says the University of California cognitive scientist, "are designed to optimize symbolic processing, but they are not sufficiently fast or efficient. Supercomputers of a different architectural design seem required for such problems."

What is needed, many computer and cognitive scientists believe, are powerful concurrent, or parallel, architectures. Says Thomas Rindfleisch, of Stanford's Heuristic Programming Project, "Highly concurrent symbolic processing may change the way we look at computers and push the entire field to a new level, enabling us for the first time to perform complex operations such as vision, or common-sense reasoning, or just getting a hunch." •

A Picture is Worth . . .

by William Check

Sometimes images are the only way a problem or a solution can be represented. Getting images into and out of computing systems is a constantly advancing frontier.

Strides—and needs—in obtaining three-dimensional images do not respect disciplinary lines. Imaging problems that are shared by research disciplines range from unraveling the composition of the earth's interior to uncovering metabolic activity in the brain, from discovering the shape of the internal organs in humans to revealing the structure of the minute elements that make up living cells. What ties these diverse areas of research together is that they all depend on a family of computer algorithms for reconstructing three-dimensional images from radiation. The radiation can be X rays, radio-frequency waves, electrons, positrons, or seismic waves.

The use of these computerized algorithms for reconstructing three-dimensional images from two-dimensional projections is a recent phenomenon in many fields. The solution to the underlying reconstruction problem, however, is not. It builds on a method developed in 1917 by Austrian mathematician Johann Radon and rediscovered repeatedly since then by mathematicians, astronomers, electron microscopists, and radiologists.

One of the computer reconstruction algorithms—used in medical diagnosis for more than a decade—has roots of its own that go back at least two decades. It is essentially the computerized mathematical algorithm for which physicist Allan Cormack and electrical engineer Godfrey Hounsfield were awarded a Nobel Prize in 1979. Hounsfield employed the algorithm in the original CAT (computerized axial tomography, now CT) scan. He built on mathematics published by Cormack in 1964 and on an algorithm called ART, for algebraic reconstruction technique, introduced in 1970 by three researchers at the University of Buffalo: Richard Gordon (now at the University of Manitoba), Gabor Herman (now at the University of Pennsylvania) and Robert Bender.

Clayton. 3-D maps of the earth's mantle.

Hager. Translating velocity anomalies.

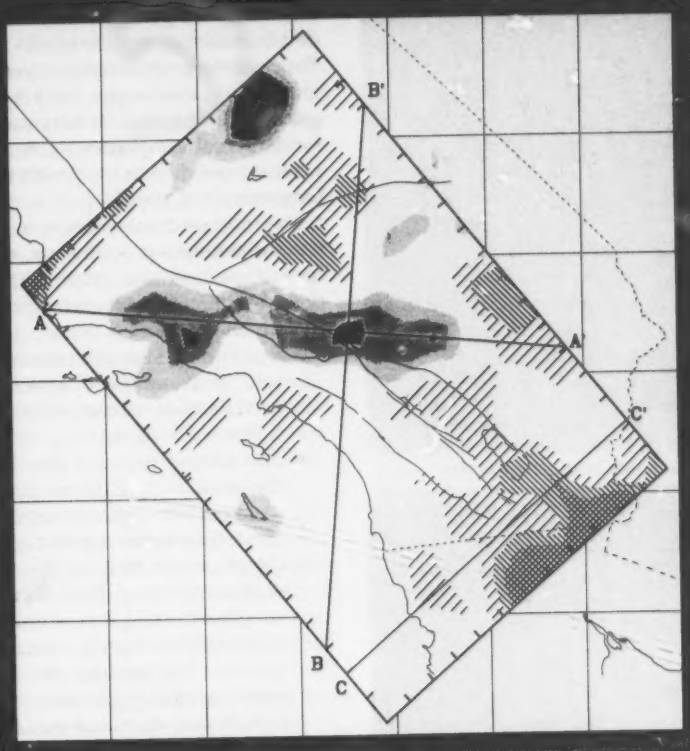
Another major computer reconstruction algorithm was pioneered in the late 1960s by a team of X-ray crystallographers led by Aaron Klug, now a Nobel laureate at Cambridge University, and David DeRosier, of Brandeis University. It is currently being used to create three-dimensional representations from electron-microscopic analyses as well as from more advanced body-imaging techniques.

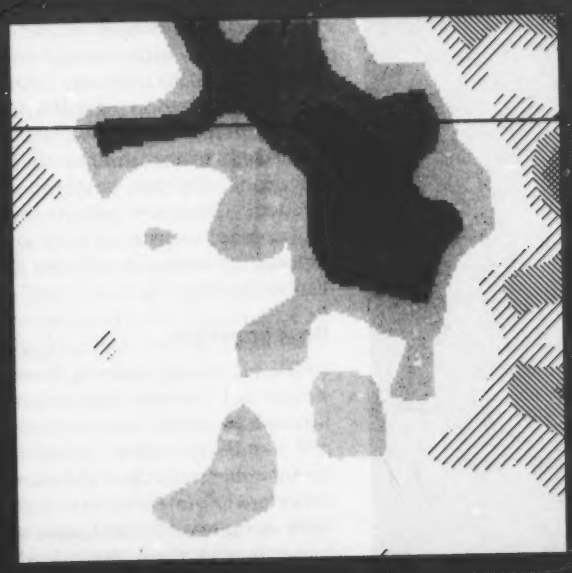
The revolution begun in 1973 by the introduction of the CT scanner is still going on, and scientists in many fields are devising ways in which computer algorithms can advance their visualizations. In each field of research, scientists are developing complicated programs that will provide new information or new ways to combine existing data.

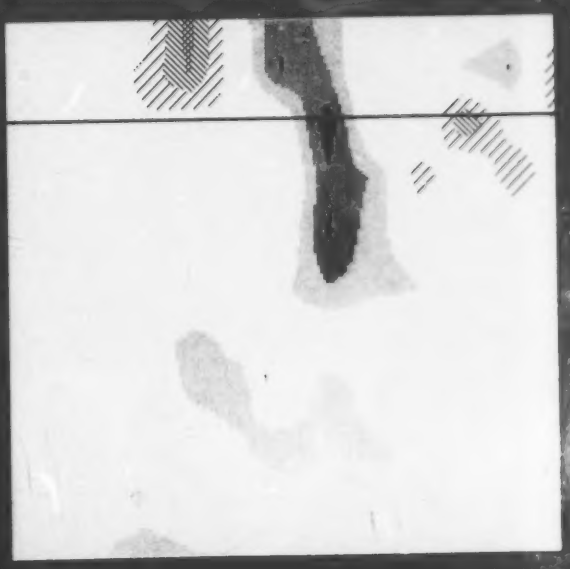
Geophysics

The classic cross-section of the earth—crust, upper and lower mantle, molten outer core, solid inner core—is the grossest average characterization of the subsurface earth. There is heterogeneity, however, reaching finer and finer scales, to be found at every level in the earth's interior.

Until recently, it was not possible to locate heterogeneity beneath the crust. But seismic energy, generated by earthquakes, is only another kind of wave-borne energy, with the earth's interior acting as the transmission medium for the seismic waves. With the application to the waves of the algorithm for three-dimensional reconstruction that was developed for CT scanning, the technique


B'
A
A'
C'
B
C





Herman. The action is in instrumentation control. Below, asymmetry in a tumored brain.

of seismic tomography was born. This tool now is providing fine-scale information about the earth's interior processes. (See "Windows to the Earth" by William J. Cromie, *Mosaic* Volume 15 Number 6.)

At the Seismological Laboratory of the California Institute of Technology, Robert W. Clayton and his colleagues are constructing the first detailed three-dimensional maps of the earth's mantle. Their starting information consists of the seismic waves generated by earthquakes. Their goal is to determine how fundamental properties such as temperature and density vary with latitude, longitude, and depth.

The reconstruction problem poses different challenges in seismic tomography than in medical tomography. There are three additional complications when the earth is the patient:

- Even in the simplest case, the seismic waves, or rays, generated by earthquakes curve as they pass through the earth.
- The seismic rays are also bent by the regions of high or low temperature, the regions that the seismic tomographer wants to find.
- Geophysicists have to live with irregular sampling. The earth cannot be placed in a CT scanner and monitored uniformly around its perimeter. The earthquake sources come randomly in time, and the signals are detected nearly randomly in space, wherever there is a seismic detection station. Moreover, these stations are set up only on land, which covers a small part of the earth's surface.

Special correction algorithms have to be incorporated into the reconstruction procedure to account for these complications. Others, however, are not needed. Clayton's patient does not move; geophysicists need not apply additional corrections to remove that particular blurring.

Back projection

Seismic tomography is based on a simple fact: Seismic waves travel faster through materials that are more rigid and less compressible. In general, this means colder regions of the mantle. Hot areas are more fluid, less rigid, and more compressible, and seismic waves travel through them more slowly.

Earthquakes generate many kinds of waves. Some skim along the surface of

the earth. Others dive into the mantle. Each kind travels a predictable route with a known velocity and arrives at a seismic station in a given sequence. Given the time of the earthquake and its distance from a detection station, seismologists can calculate the time it should take each kind of wave to arrive at that station. When a wave takes either more or less than the predicted time to travel to the station, it is assumed that the wave has passed through a travel-time anomaly, a region of altered density and temperature in the earth's mantle.

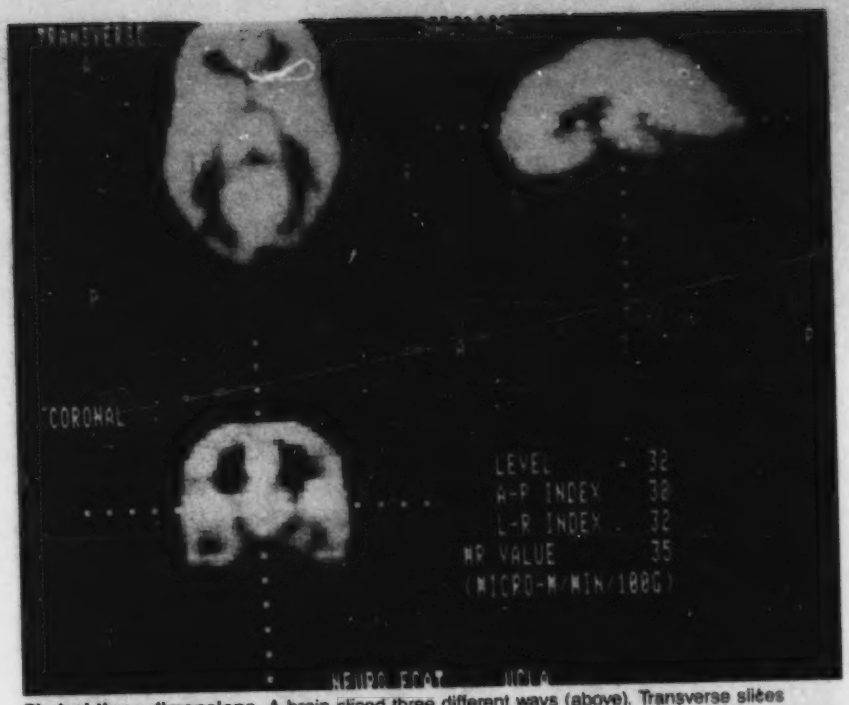
Seismologists can draw a line between the earthquake and the detection station, and they know that the anomalous mass lies somewhere along that ray. However, they cannot tell from a single observation where along the line the anomaly lies. It is here in the analysis that the computerized reconstruction algorithms are brought into play to locate as precisely as possible, in three dimensions, the position of the hot or cold spot. This method, called algebraic back-projection, is very similar to the algorithm first used in CT scanning.

In algebraic back-projection, the anomalous observations are distributed along the travel path of the earthquake wave. The mantle is divided into a three-dimensional grid of unit cells, and a unit value representing a discrete piece of the anomaly is entered into each unit cell lying along the path of the anomalous ray.

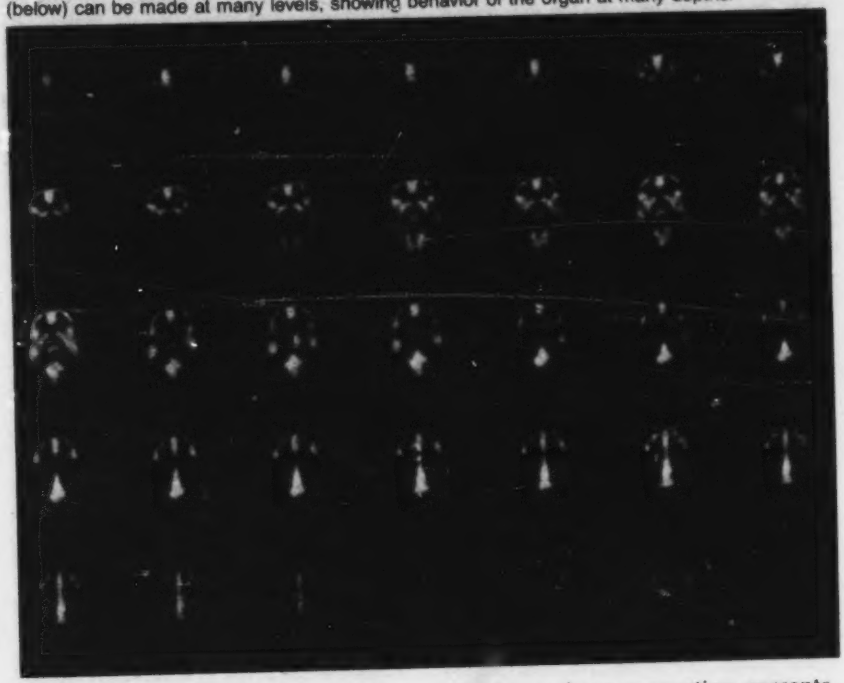
In many cases, more than one ray will pass through an anomaly. It is difficult for scientists looking at two intersecting waves, each producing an anomalous observation, to tell whether one anomaly contributed to both waves or whether the two passed through independent anomalous zones. The computer, using the back-projection algorithm, can solve this problem.

Essentially a velocity is assigned to each unit cell. The computer tentatively solves a large number of simultaneous equations in which the velocities of all unit cells through which each wave passes must produce the net resultant travel time of that wave.

Because the anomaly was distributed along the entire length of the line, the initial computer map of each anomalous mass will be smeared over a larger area than the mass actually occupies. This blurred image can be largely removed by re-solving the simultaneous equations several times in such a way that the anomaly becomes concentrated in a smaller and smaller volume, in fewer and fewer unit cells.



Phelps' three dimensions. A brain sliced three different ways (above). Transverse slices (below) can be made at many levels, showing behavior of the organ at many depths.



Convection currents

Several novel and important features of the earth's mantle have been discovered through use of this procedure. Some of the most important information has been about convection currents. Previously, it was known that there are hot and cold regions in the earth's upper mantle. Geophysicists surmised that the lighter, hot regions are rising and the denser, cold regions are sinking.

Seismic tomography has revealed that beneath hot spots, volcanic provinces at the earth's surface, hot material extends

deep into the lower mantle. In other regions, the lower mantle is generally cold beneath places where subduction, which provides volcanic provinces, occurred some 100 million years ago. This suggests a rather sluggish circulation, much lower than the velocities of the tectonic plates at the surface of the earth. (See "What Drives the Earth's Plates" by Patrick Young, *Mosaic* Volume 10 Number 5 and "When North Becomes South" by Derral Mulholland, *Mosaic* Volume 13 Number 5.)

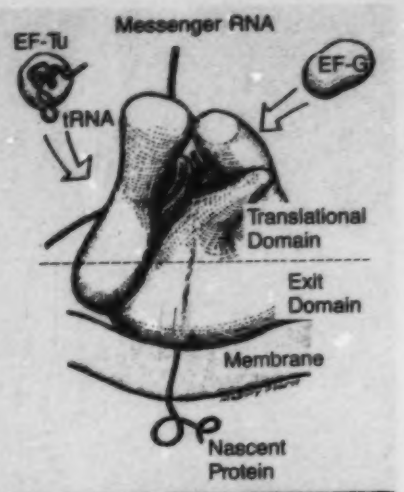
These initial successes of seismic tomography have emphasized the need for more advanced computer analyses in geophysics. The next challenge, says Robert Clayton, is to confirm the hypotheses arising from the method.

Bradford Hager at Caltech is building computer models that translate the velocity anomalies into regions of altered density, because the high-density regions that increase the velocity of seismic waves also affect the intensity of gravity. "It's been known for some time that the earth's gravity field is not that of a sphere," Hager says. The goal of the models is to predict the irregularities in the earth's gravitational field based on the subsurface density variations.

A simple integration over all subsurface anomalies to obtain the net surface gravity gives the wrong answer. The problem, Hager explains, is that on a time scale of millions of years, the earth is not rock, but a viscous fluid. Dense regions not only sink but drag adjacent material along with them and push underlying material out of the way. This results in dynamic deformation of the earth's surface and interior density boundaries that also affect the gravity field. When they construct more complicated algorithms to take this flow-induced topography into account as a function of the mantle's temperature, Hager and Clayton obtain a much more satisfactory map of gravity distribution. This model, however, includes the effects of variations in viscosity as a function of the mantle's temperature. Refining the model demands even more detailed analyses, says Clayton, "and that's going to require a lot of computing."

Clayton and his colleagues are also attempting to stretch seismic tomography in another direction. They believe it can be used to locate hydrocarbon deposits within the earth's upper crust, kilometers below the surface. Reflected acoustic waves reveal these deposits as humps in the surrounding geologic layers. Correct interpretation of the reflections, however, demands more accurate knowledge than is now available of the velocity of the acoustic signals as they pass through the various layers.

Clayton's group has been working on this problem with a computer model similar to the one they used for seismic tomography of the deeper layers. They can control these acoustic signals and the detection devices better than they could with signals from earthquakes. To evaluate this process, they are constructing a more complicated computer model based on integral equations called Fourier transforms. It is the same type of model that is used in modern tomographic medical imaging.



Ribosome domains. James Lake's model of an extended chain of nascent protein unfolds as it passes through the ribosome.

Reconstruction problems

In another type of tomography, called PET, or positron emission tomography, the computer algorithms assist in the analysis of biochemical activity in various regions inside the body, chiefly the head. Says Michael Phelps of the University of California at Los Angeles, co-designer of one of the first PET scanners, "With PET we have the ability to address directly probably one of the most critical topics in medicine, the biochemical basis of disease and therapy." PET is also revealing a body of fundamental information about the localization of mental functions, such as memory, speech, and even thinking, in various brain regions. (See "Seeing the Brain at Work" by David Zimmerman, *Mosaic* Volume 12 Number 3.)

Formally, the task of obtaining a three-dimensional image from a PET scan shares some features with seismic tomography. In geoscience the signal is generated by an earthquake. In PET the signal comes from the decay of a positron-emitting isotope that has been taken up by a cell. In both cases a signal detected outside the volume of interest must be traced back, to identify its origin within the volume.

There is a major difference between the two systems. In seismic tomography the location of the earthquake is known; it is a spot between the event and the detector that must be located. In PET scanning, it is the location within the body of the radioactive decay that must be determined. When this is done for a large number of disintegrations, the data can be plotted as a map of the distribution of the radioactive isotope throughout the brain. Because the isotopes are coupled to compounds that are involved in the body's energy metabolism, such as glucose, oxygen, amino acids, and fats, the map shows the level of metabolic activity in the different body regions.

The primary data in PET scanning of the brain are obtained by using crystals surrounding the head to detect the photons from the radioactive decays. Each positron travels a very short distance and then collides with an electron and generates two gamma-ray photons at an angle of almost 180 degrees to each other. When two detectors on opposite sides of the head each register a photon

---

Check writes on biology and medicine.

within a small time interval, the computer records a positron emission.

The source of any positron lies somewhere within a well-defined region on a line between two detectors. The computer assigns all the events occurring between any detector pair to the line between them. Finding the distribution of the isotope then requires solving a series of integral equations representing each detector pair.

Because of its unique properties, PET scanning presents additional challenges not encountered in CT scanning and seismic tomography. Each of these problems requires new algorithms or more computing power, or both.

For instance, the amount of the isotope-bearing compound in any given brain region may be changing during the assay. Says Phelps, "In PET scanning we are not looking at static anatomy, as in CT. We are looking at a dynamically changing biochemical system. So the computer needs to understand biochemical kinetics. This adds computational complexity." (See "Resurrecting the Ignorabimus" by George Alexander, *Mosaic* Volume 14 Number 3; and "Solving for the Unsolvable" by Henry Simmons, *Mosaic* Volume 10 Number 1.)

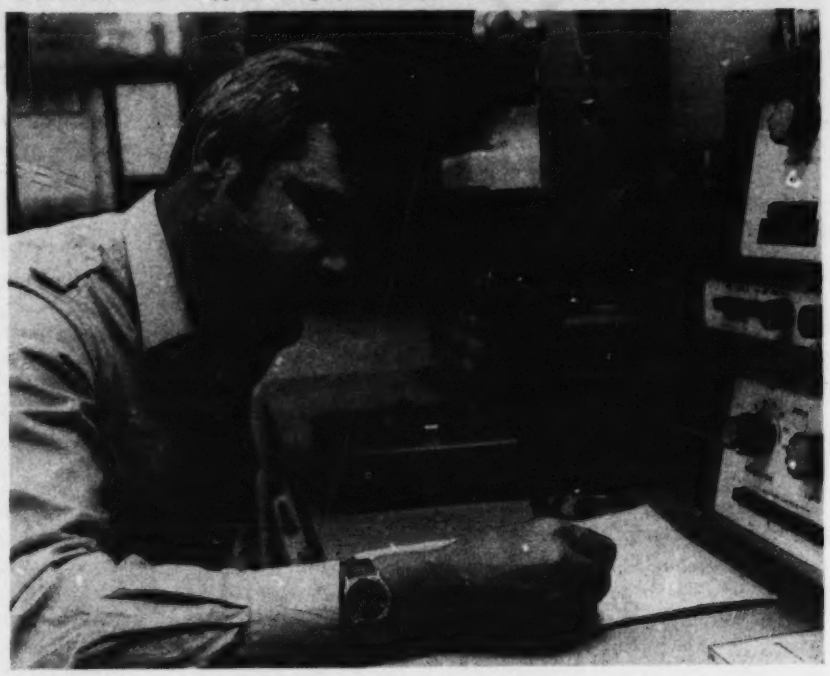
Currently, the favored approach is to obtain isotope distribution images at several time points and then combine the biochemical values into a kinetic model. Phelps believes this is far from the best solution. "If we were bold enough," he says, "we would deal with this as a reconstruction problem that simultaneously varies in both space and time. That's a computational problem that will scare the pants off you."

The ideal way to approach this situation, Phelps says, would be by way of a branch of mathematics called estimation theory. Essentially this method asks, What is the most likely reconstruction of the isotope variation in space and time? This method would call on mathematical procedures such as least-squares estimates or maximum-likelihood (maximum entropy) computation.

Erasing a blur

According to Phelps, a maximum-likelihood estimation has been fairly well worked out by several people for the spatial distribution component alone. Factoring in the kinetic component, though, increases the size of the data base and the complexity of the problem. "Currently some people at MIT are trying to do this using a *Cray* and a fast gigabyte laser disk," says Phelps.

Aebi. Electron microscopy and image processing reconstruct noncrystalline actin.



The display of the mass of data collected during a series of PET scans in a way that an observer can interpret is also a problem. "Nobody yet knows how to do this satisfactorily," says the UCLA scientist.

He and his co-workers at UCLA are developing a large, buffered video display with a rapid-refresh memory. The display screen shows the head in three perpendicular sections, coronal, sagittal, and transverse. Isotope intensity is represented by shades of gray and white.

The observer can instruct the program to call up sequential views of the brain in any plane. When the brain is leafed through coronally, the other two views show what level of the brain each coronal section comes from. At the same time a fourth representation

shows numerical values of the isotope throughout the coronal section.

The UCLA scientists are tackling simultaneously the problem of improving the initial data collection. The blurring of data during back projection in PET scanning is especially pronounced because of the two-step nature of the detection process. Locating the site of photon production by the positron-electron annihilation does not give the position of the positron emission. The positron travels a minute distance before it encounters an electron—enough to decrease resolution.

To improve resolution, Phelps and his colleagues have turned to a method called signal amplification. It consists simply of having a much greater number of smaller detectors. Rather than 10 or 20 photon detectors, Phelps and his colleagues are now testing a PET scanner with 1,024 detectors. Each detector registers fewer events, but each event can be traced to a smaller-volume element in space. Because the detectors occupy three parallel image planes, the total number of possible detector pairs is 64,000, two orders of magnitude greater than current PET scanners.

This new scanner has required new hardware. Instead of the standard array processor used to perform the Fourier transforms in CT scanning, the advanced PET scanner requires a hard-wired sorter. In addition, its information-gathering ability is pushing available storage capacity. "We would rapidly fill a 500-megabyte disk," says Phelps. "We had to install a multigigabyte storage system."

Much further expansion of this imaging instrument's data-generating ability threatens to outrun current hardware.

Controlling the scanner

The most recent addition to the family of imaging instruments is the NMR (nuclear magnetic resonance) scanner. Like PET, it is still in its developmental stage, although in wider use. Scanning with NMR depends on the propensity of protons, hydrogen nuclei, to align their magnetic spins when subjected to a very strong magnetic field. Hydrogen nuclei are present in all nucleic acids, proteins, and other organic compounds. The patient or sample is placed within

Architecture and algorithms

Computer algorithms turn an integrated distribution of x-ray density, positron emission, electron patterns, or radio frequency waves into three-dimensional images by solving a large system of integral equations. Solving these equations takes considerable time, even for a computer. However, emerging improvements in computer design, coupled with adaptations of the software used for image reconstruction, are decreasing the time needed to run these programs.

One source of improvements in image reconstruction is the emerging change in the way computers are put together, their architecture. This term refers not to the macroassembly of the computer components but to the internal construction of the processors that carry out the machine's operations. Improvements in computer hardware will spring from the widely sought innovation called parallel processing, in which processors will be given the ability to perform several steps at once and then combine the partial results to give the answer the user wants. (See "The Pursuit of Parallelism" by Margaret L. Silbar, *Mosaic* Volume 16 Number 3.)

In traditional computers, "there is only one kind of relationship between the units of computation," says Jim Browne, of the University of Texas, Austin. "That's a sequential relationship." For every basic operation routinely carried out, parallel processors are designed to break the operation into several independent steps, execute the steps, then combine them to give a result. For instance, a floating-point multiplication may be broken down into 220 steps through a combination of microcode instructions and hard-wired logic in the processor. A graph of the resulting computation is not a straight line, but a two-dimensional map, as variables get shifted among operations that are going on simultaneously.

Such processors can also more quickly handle what computer designers call vectors, matrices of numbers that represent values to be operated on in the program. The processors can add or multiply two entire vectors and do it quickly because of their ability to store the vectors in preselected locations so that one element can be fetched from memory while another is undergoing the operation. The result is that a single hardware operation produces multiple results by operating on several variables at one time. The approach is called pipelining, because one computation is fed into the pipeline (the processor stream) before the previous computation is over.

This type of parallelism, internal parallelism, is inherent in the processor's instructions but invisible to the programmer. A number of recent supercomputers incorporate this approach. It goes only part of the way, however, to a truly parallel architecture.

True parallelism, on the other hand, combines internal parallelism with parallelism of a higher order. In true parallel processing, several processors are controlled by the software, enabling them to perform simultaneous computations on a problem.

Images

In most conventional image-processing algorithms, the first step is to determine the range of values of gray levels in the picture elements, or pixels. This is done by setting up a histogram of the number of pixels of each value. The analysis is performed on a two-dimensional array of pixels, one column at a time. Then values from two columns are added as pairs. This is a neatly defined set of steps in which the addition of two columns does not depend on the result obtained for either column.

Image-reconstruction algorithms especially lend themselves to solution by parallel techniques, says Jim Browne. "The dependencies are predictable and the computations are regular." Which operations can be carried out independently of each other, and thus simultaneously, can easily be determined.

It is easy to design dedicated computers to carry out these tasks, Browne says. "In fact, you can even match the machine to the program. If you build a certain number of signal channels into an NMR scanner, you can build that same number of channels into the processor."

the field, and the nuclear spins align in response. The introduction of electromagnetic energy in radio frequencies causes the nuclear spins to resume their random distribution, a process called relaxation. A signal given off by the nuclei as they relax—responding to an array of electromagnetic frequencies—provides information about the protons' microenvironment, which can be used to form very detailed images of the interior of the target.

In the view of mathematician Gabor Herman, who for many years has been involved in writing reconstruction algorithms for CT scanners, "the action right now in NMR is not in the mathematics of reconstruction, it's in the control of the instrumentation." Herman is referring to the way in which the magnetic field can be modified during imaging to improve the data from which body regions can be reconstructed.

Typical modifications include the creation of a magnetic field gradient on which the radio frequency pulses are superimposed.

Use of complex, time-varying magnetic fields and field gradients with radio frequency pulses can generate relaxation signals that carry far richer information. For instance, the correctly shaped magnetic field gradient can produce a signal from a very restricted part of the space being imaged. A series of such gradient pulses gives information about points throughout the space. Such techniques yield data that require very little mathematical processing.

To obtain this information, the NMR scanner must be controlled by computer programs that turn the correct combination of electromagnetic coils on and off in fractions of seconds and in the correct sequences. These computer programs are still being developed. No single approach is yet dominant, says Herman. "It's still a matter of great dispute."

A second challenge, which Herman is now working on, is to use the stored reconstruction data to display selected organs in the body. He has, for instance, created isolated images of the ventricular system of the brain and the ventricle of a beating heart.

This is accomplished by computer programs that operate on the total three-

Browne's specialty is parallelism in hardware and systems. Paul Swarztrauber of the National Center for Atmospheric Research in Boulder, Colorado, works in the complementary discipline of algorithm design.

"The biggest worry in the software field in the last five years has been adapting the fast Fourier transform and other algorithms to the new vector, pipeline, and parallel computers," says Swarztrauber. "For increasing the speed of computers, it is almost as important to have these new algorithms as the new hardware." When a traditional FFT, for fast Fourier transform, algorithm is run on a new machine, the actual computational speed is only one-fifth to one-half the potential gain. When the program is reformulated to take advantage of the new hardware, the potential speed increase is realized. (See "Algorithms" by Arthur Fisher, *Mosaic* Volume 16 Number 3.)

The key phrase, Swarztrauber says, is converting sequential computations to independent computations. One of his publications carries the title "Vectorizing the Fast Fourier Transform."

Vectorizing the FFT

As a simple example of making a sequential problem independent, Swarztrauber describes the addition of as many as 100 numbers. On a traditional computer, this would involve setting up a loop. The computer would keep a running sum, S_i, in storage and would serially solve the equation $S_i = S_{i-1} + X_i$, where X_i is the next number to add. With this approach, the computations are not independent. For example, S_{20} cannot be computed until S_{19} has been. Thus, this calculation is sequential.

In the modified computation, the numbers are first added as pairs and then these sums are added, and so on, until the sum of all the X_i entities is obtained. Because the additions are independent, they can be computed simultaneously, or in parallel.

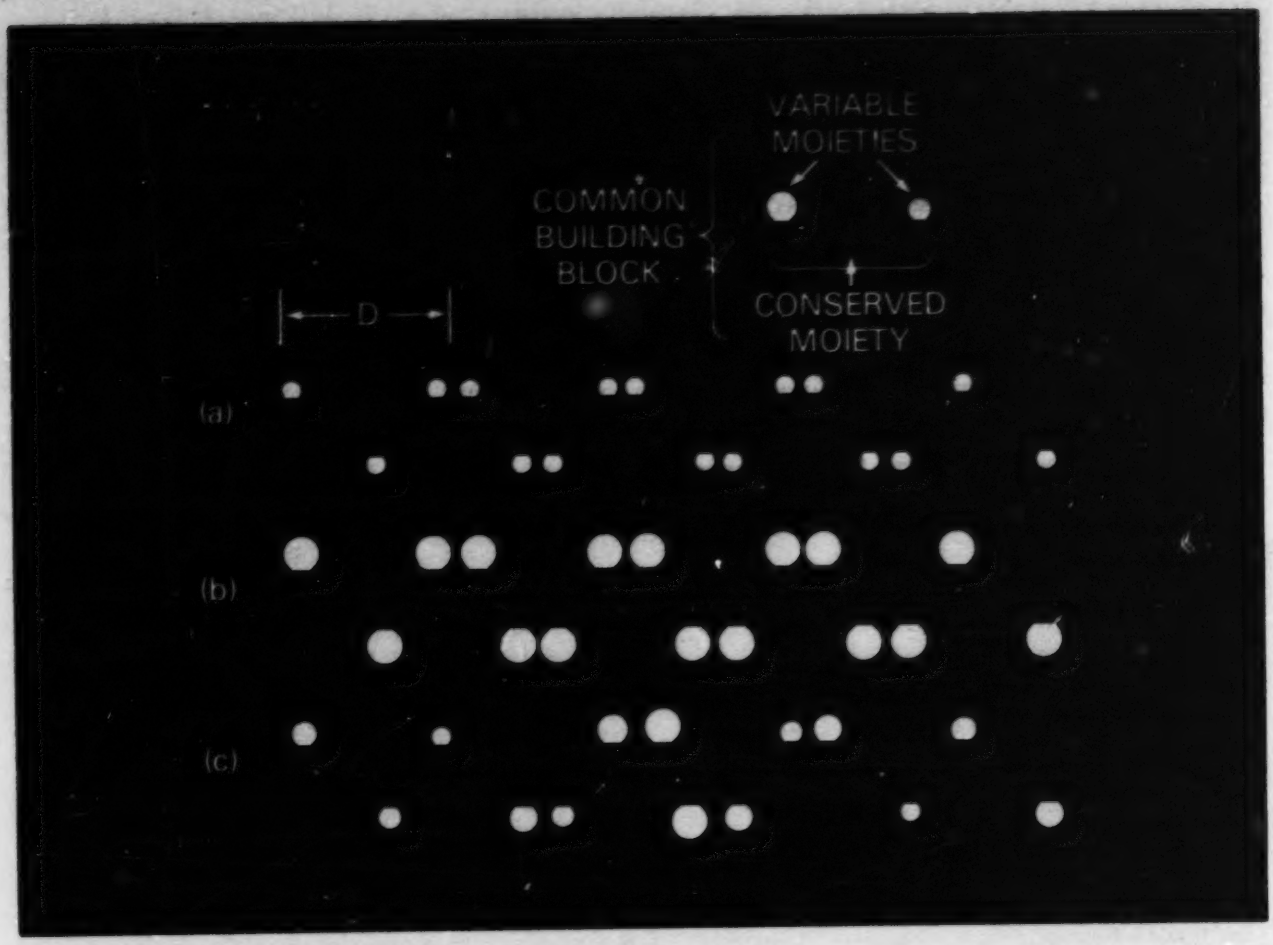
"The fast Fourier transform is really a matrix-vector multiply," Swartztrauber says. "However, because of the very special form of the matrix, it can be done in fewer than N^2 operations, where N is the order of the matrix. In fact, it can be done in N log N operations. The FFT is a good example of a computation that requires some work in order to adapt it to the new architectures."

Two of Swarztrauber's algorithm modifications that take advantage of pipeline processing are lengthening the string and loop inversion. Lengthening the string refers to increasing the number of computations in the pipeline at any one time. Efficient computations in a pipeline computer are done in an assembly-line manner. The efficiency is largely determined by how long it takes to start the assembly line. That is, how many operations over how much time does the processor have to do to crank out the first result. This is the start-up time.

If there are 100 computations coming down the pipeline, for instance, for a particular program, the pipeline processor might turn out its first result on the fifth operation and produce a result on each operation after that for 100 operations. As Swarztrauber says, "The start-up time is amortized over 100 calculations, so you want to make sure that your vectors, numbers of computations, are as big as possible." This feature is built into the modified software.

Loop inversion provides a solution to a problem inherent in the FFT. "It's in the nature of the fast Fourier transform that the vectors tend to get shorter and shorter as the calculations proceed," Swarztrauber says. With the loop-inversion feature, the number of computations on the line decreases for a while but then increases. This feature alone speeds up the FFT twofold on pipeline machines, the NCAR scientist declares.

At the present time, says Swarztrauber, hardware designers are building parallel architectures that "are efficient for many computations but not all. The most exciting development right now in computers is the trend for people in architecture and algorithms to get together in the design of computers. Then algorithmists like myself won't have to do so much work adapting algorithm to hardware; the net result will be a faster computation." ●



dimensional image data. First, a program identifies the volume elements that belong to the structure of interest. Because the computer program can select a part of the volume data, the final picture can show either a whole organ or a dissection in any plane. Then a three-dimensional surface is constructed. (Mathematically, this operation consists of a description of a set of polygons.) Finally, computer graphics packages are used to display the surface of the organ.

Building an Image

At a much finer spatial level, electron microscopists are using three-dimensional reconstruction algorithms to obtain images of supramolecular assemblies that make up living cells. The programs used by these structural biologists are usually based on application of the Fourier transforms as formulated by Aaron Klug and David DeRosier.

James Lake of UCLA began using these techniques in their initial form to probe

Steven. Different proteins polymerize into filament structures derived from electron microscopy and amino acid sequence.

the structure of the ribosome, the cell's protein-synthesizing factory. He explains that electron microscopy can yield overall sizes and shapes for objects not amenable to study by X-ray crystallography. X-ray crystallography requires large, three-dimensional crystals. Some cellular components (such as the ribosome) readily form the large crystals needed, but some proteins do not. Because an electron beam can be focused by electromagnetic lenses, an image made in an electron microscope has both amplitude and phase information about the diffracted electrons. An X-ray diffraction pattern contains only amplitude data. So a repeating lattice is not necessary to obtain structural information with electron microscopy. A disadvantage of electron microscopy is that its resolution for biological specimens is only about 10 to 20 angstroms, much less than X-ray crystallography.

Over the past ten years Lake has defined the ribosome's contours well enough to find the ways in which large

and small subunits fit together and in which the ribosome binds many of the components for protein synthesis—messenger RNA, transfer RNAs, amino acid protein components, and the like.

At Johns Hopkins University's medical school, Ueli Aebi is using high-resolution electron microscopy and image processing and reconstruction methods to analyze the three-dimensional molecular architecture of actin and intermediate filaments, two of the basic constituents of the cytoskeleton found in most eucaryotic cells. Actin in particular is an ideal candidate for electron microscopic rather than X-ray crystallographic study. Although actin is a protein, it has not been crystallized in three dimensions. It does, however, readily form two-dimensional crystals. (See "Scaffold of the Cell" by Edward Edelson, *Mosaic* Volume 12 Number 4; and "Ribosomal RNA" by David Holzman, *Mosaic* Volume 16 Number 2.)

Actin filaments are the important elements responsible for generating force in a variety of contractile processes (including muscle contraction and cellular motility) and in the maintenance and dynamics of cell shape and cytoplasmic consistency. In attempts to understand the molecular basis of these action-based motile processes, Aebi strives to determine a three-dimensional model of the double-helical actin filament. This will be a model in which the shape and relative orientation of the individual actin subunits can be resolved and in which the binding sites for contractile, regulatory, and cytoskeletal proteins, which specifically interact with actin, can be mapped.

One approach being taken by Aebi is to use computational fitting procedures and model building to align three-dimensional models of the actin subunit within actin filament reconstructions. For this purpose he has determined a three-dimensional model of the actin subunit from a set of electron micrographs recorded from two-dimensional crystals tilted at various angles to the electron beam in the instrument. To arrive at a near-atomic model of the actin filament, he plans to fit emerging X-ray models of the actin molecule into filament reconstructions.

Aebi notes that one of the advantages of the Fourier methods that he uses is their transparency with regard to the steps leading from a set of two-dimensional projections to a three-dimensional

Swarztrauber. Sequential into independent.

reconstruction. The major disadvantages of Fourier methods are their relatively slow computing speed and the difficulty in applying them to noncrystalline and nonsymmetric specimens. It was to overcome these limitations, as well as some of the disadvantages inherent in several other reconstruction methods, that Gordon, Bender, and Herman devised ART.

ART was first applied to the three-dimensional reconstruction of individual ribosomes from electron micrographs of negatively stained specimens, and modification by Hounsfield for his CAT scanner followed. More recently, other modified versions of ART have been implemented by several structural biologists trying to obtain three-dimensional models of large nonsymmetric objects, such as ribosomes, cell organelles, and even whole cells.

Intermediate filaments

In addition to the role played by actin, a cell's morphology is maintained by other filaments. These include the intermediate filaments, whose structures are studied through electron microscopy by Alasdair Steven of the National Institutes of Health.

The proteins that make up intermediate filaments fall into five classes of 25 to 30 subunits each. This diversity gives rise to a fundamental structural problem, says Steven. "How do all these different proteins polymerize into basically the same filament structure?"

Steven is approaching this question by using a STEM, a scanning transmission electron microscope, to define the structure of various kinds of intermediate filaments. The STEM can provide good images without the use of heavy metal stains, which are necessary for other electron microscopy but which can lead to ambiguities of interpretation. Application of a reconstruction algorithm to the STEM data has contributed to a general structural model that agrees well with the implications of the first complete amino acid sequence for a keratin subunit, which Peter M. Steinert of the National Cancer Institute published in 1984.

Steven believes that further improvements in the resolution of three-dimensional images obtained from electron microscopy do not depend as much on further refinement of reconstruction algorithms as they do on progress in preparing the specimens so as to preserve native structure, and in recording electron images at tolerably low electron doses to diminish radiation damage. Computationally, the challenge lies in developing methods to reduce the noise levels of the low-dose images whose signal contents contain the high-resolution information. "There is real momentum in this area," Steven says.

Highly crystalline specimens may be treated by Fourier transform methods, but, as Steven says, "Fourier transforms can't handle disordering well." Real-space methods are under intensive development to get them to perform the averaging of many images of ostensibly the same specimen in order to suppress the statistical noise.

These methods are called correlation averaging and correspondence analysis. They work by aligning all the noisy individual images, bringing them into precise spatial register with a reference object and hence with each other. They may then be averaged so that the random noise is suppressed and the consistent signal component is reinforced. This work is only a few years old and has involved contributions from a number of groups worldwide. When these programs are fully mature, Steven says, "they will bring a great improvement in images and a much wider class of specimens that we can tackle at high resolution." •

The Problem of Macrovariables

by Edward Edelson

Many economists, like their colleagues in the ocean, earth, and atmospheric sciences, deal with systems so massively fluid as to defy—so far—efforts to model them precisely.

Like all social sciences, economics is the subject of an unending debate: How much of it is social and how much is science? The use of computers as a central tool of economics has given this debate sharper edges. By allowing economists to manipulate vast quantities of data, computers have made economics more quantitative. The debate now centers on the degree to which computer-based economics can match the complexity of the real world.

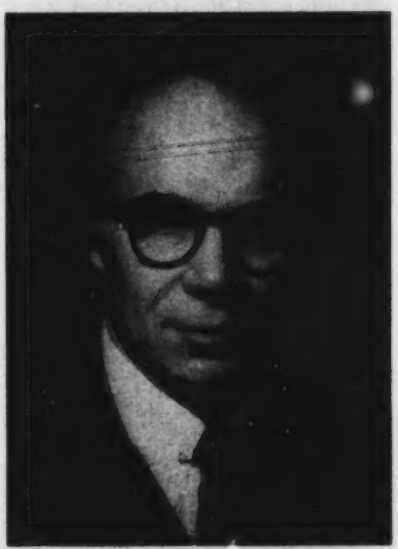
Unquestionably computers have transformed the way most economics is done, both in research and application. On the practical level, there is a thriving business in computerized data bases of detailed statistics about the economy. There is an equally thriving field of macroeconometrics, in which data are used in computerized models to predict the future of that economy.

Forecasting is part of macroeconomics, or the realm of the national and world economies. One tool of macroeconomics is simulation, in which computer-based models are used to evaluate the effects of basic economic policy changes. Computer simulations are also used at the other end of the spectrum: microeconomics, the study of markets and of individual economic behavior. Here, economists use computer programs to model the behavior of individual activities such as job-hunting and job-leaving. There is another small but growing field: experimental economics. Computers are being used in experiments that duplicate, on a small scale, such essential or representative economic activities as auctions.

Forecasts

The use of econometric models in forecasting is both the most prominent and the most controversial activity of computer-based economics. Forecasts based on dozens of econometric models are routinely used by corporations and governments to plan strategies.

But different models can produce widely differing predictions, and economists hold equally differing opinions about their worth. Proponents such as Lawrence Klein, of the Wharton School of Economics at the University of Pennsylvania and one of the leading figures in econometrics, say that the accuracy of predictions using econometric models has improved markedly and that these models have proved themselves as practical tools. Others disagree.



Klein. Improved accuracy of prediction.

"Many economists believe that these forecasts are worthless," says Dale Jorgenson of Harvard University. "If you look at the big events—the oil crisis, monetarism, Reaganomics—forecasts have given no useful information on where the economy was going or when it was making a big change." Klein acknowledges that economists did not predict such major changes in the world economy. "But once they occurred, the forecasts of their effects have been very good," he says.

It is easy to explain the feeling that forecasting is—and probably always will be—an imperfect science, says Klein. One reason is that it is based on imperfect data, both for macroeconomics, or events in the economy as a whole, and for microeconomics, or pictures of individual behavior.

Macroeconomic statistics are never as good as they should be, says Klein. Data fed into economic models usually are revised, often substantially, as new information comes in. "We don't know what's happening, so how accurately can you know what is going to happen?" he asks. "We never know the true gross national product, the true investment level, and true employment level.

"First, we don't know how many people there are, so we don't know what their income is. And at the micro level, the data are as noisy as they are at the macro level. Frankly, at the micro level people don't even know for sure what they've done. Given those errors, it's a wonder we produce any useful statements at all."

The point about the forecasts produced from econometric models is that they are just as imperfect and just as useful as weather forecasts, says Klein, who won the Nobel Memorial Prize in Economic Science in 1975 for his achievements in economic modeling.

He has been working in this field for some four decades now. "We're many times more accurate now than we ever thought we would be," he summarizes. He considers a forecast to do well "if its error band is no bigger than the error band of the data." By that standard, Klein says, economic forecasting must be considered a success.

Klein and Otto Eckstein, a Harvard economist who died in 1984, played major roles in making it so—and, incidentally, in producing a major commercial product in the form of computerized data bases. Klein's work on economic forecasting began at the University of Chicago in the 1940s. The breakthrough, he says, came in the early 1960s, when sufficiently powerful computers became available.

Klein and his colleagues developed a computerized model of the economy in the early 1960s. Eckstein then developed a time-sharing system that allowed widespread use of such models, and in the late 1960s persuaded Klein to distribute his model on the time-sharing system. The model distributed by DRI, Inc., the company Eckstein founded, was a combination of Eckstein's and Klein's work. It became the most widely used econometric forecasting tool.

Models

A computer model consists of a number of equations into which factors affecting economic performance are inserted—weather related to crop yields, congressional action on taxes, changes in the money supply, and so on. The model then generates data on future economic activity. To support the DRI model, an adequate data base was needed. DRI used a data base developed at the Brookings Institution. What DRI did when it made that data base available was to put the economic data that the government gathers into a form more suitable for use in economic models. That "reliable, well-indexed product," says Dale Jorgenson, soon became the commercial tail that wagged the dog. "If I ask myself where does this activity show up in the GNP, what is the big-ticket item, this data base is it," Jorgenson says.

The theoretical work leading to today's generation of economic models began well before the modern computer. The models developed by Eckstein and Klein consisted predominantly of what are called linear simultaneous equations. To give an oversimplified example, such an equation would describe the way a change in the energy tax rate would affect the demand for energy.

Several other economists, notably Ragnar Frisch of Norway, Michael Kalecki of Poland, and three British economists, John Hicks of Oxford and Richard Godwin and Nicholas Kaldor of Cambridge, developed a somewhat more complex method using what are called nonlinear relations.

A nonlinear equation recognizes a symmetrical relationship between economic variables. For example, if the energy tax is raised, the resulting change in the energy demand also affects the demand for labor. Similarly, a tax on labor affects the demand for energy. Nonlinear equations describe these simultaneous, interrelated feedback effects.

Nonlinear equations were first used in microeconomic models. One of the leading figures in the development of those models was Eugene Slutsky, a Russian statistician who published his work in an Italian economics journal in 1913. The subject was further developed in the 1930s by John Hicks and Paul Samuelson, another Nobel economics laureate, and has been of fundamental importance in microeconomics for the past three decades.

Although nonlinear equations were used in macroeconomics as early as the 1930s, their widespread use had to wait for the development of modern computers. They became especially important in the 1970s. One reason, Jorgenson explains, is that "the energy crisis showed that nonlinearities could be important." Until then, he notes, economists generally had assumed that the demand for energy was not elastic, that is, that an increase in the cost of energy would not have a major effect on the demand for energy. But when oil prices increased tenfold in the 1970s, the demand for oil dropped sharply, showing that there was a feedback relationship between energy price and energy demand. The demand for labor also rose, showing the effects of nonlinearity.

A number of econometric models are available commercially from such organizations as DRI and Chase Econometrics. These produce well-publicized forecasts of the gross national product or unemployment for the coming quarter or year. Other econometric models are used for research. Because of their complexity, none of these models would be workable without modern computers. Early models had 30 or 40 equations. Some today have as many as 15,000.

For example, a model developed by Ray C. Fair of Yale University has 128 equations for what are called endogenous variables—gross national product, interest rates, money supply, and other integral parts of economic activity. Another 110 exogenous variables describe factors outside the field of economics that affect economic behavior—population size, defense spending, and similar items.

There are two basic reasons why such econometric models don't do as well as they might, Fair says. The first is that "we have a lot to learn about the econo-

Jorgenson. Both better data and more thorough analysis of existing data are needed.

my," which has uncountable details still evading quantification. The second is that most models make the assumption that the economy is stable, that the future will be like the past. If the economy is stable, models will be good in the long run, Fair says, but if it is not stable, models are likely to be of limited use. "Whether the economy is stable is an open question," he says.

Unquestionably, unexpected events—or outliers, in the jargon of economics—keep occurring. The Depression was one such outlier, the Arab oil embargo of the 1970s another, and the deep recession that occurred at the beginning of the 1980s was yet another. Fair's model uses the assumption that in spite of those outliers, the basic structure of the economy remains the same. This assumption gives the economist a better handle on modeling, Fair says. But there is no way to test it experimentally. "We just wait and see."

Assumptions

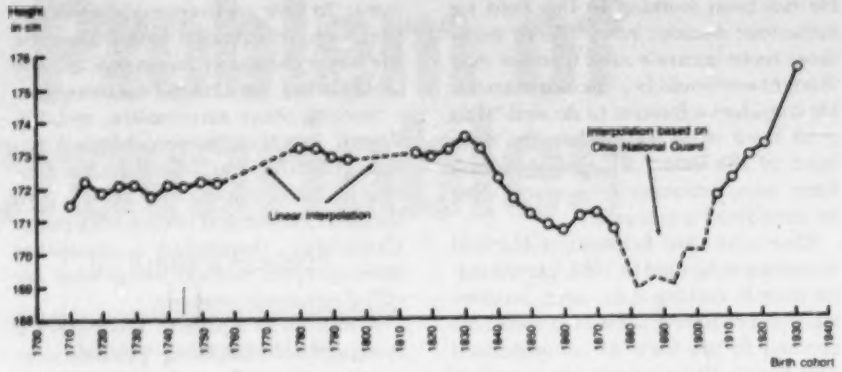
One unquestioned fact is that, in practice, commercial forecasters constantly tinker with their computer models in an effort to make their forecasts more accurate. Essentially, Fair explains, they start with an idea about the direction the economy is going. Often, the models' results disagree with their expectations. It is commonplace to change the coefficients in equations so that the results are closer to the economist's predictions, the Yale economist says.

A rationale for making such changes in the model is that, although economic theory is well-developed enough to allow fairly reliable predictions, modeling is not. The changes thus are explained as an effort to make the model fit the data more closely. Like most assumptions in economics, this one is the subject of controversy.

Two economists in Minneapolis, Christopher Sims of the University of Minnesota and Robert Litterman of the Federal Reserve Bank, have developed a countervailing theory leading to a simpler econometric model.

"Very careful theorizing leads us to believe that everything in economics is related to everything else," Sims explains. "There are just too many unknown parameters to let us fit the model to the data."

Edelson was the author of "Phylogeny Without Fossils," Mosaic Volume 14 Number 6.



Heights. White North American males began growing taller in the decade of the 1890s.

Sims and Litterman maintain that conventional forecasting has a small chance of success because the basic theories are inadequate to describe the economy's complexity. Most models contain a large degree of subjectivity, Sims maintains, because there is no reliable set of theoretical assumptions, as there are in physics or chemistry, to describe the behavior of the system.

The solution to this problem is to assume that the best guide to the economy's immediate future is its immediate past, with the addition of a trend term. Essentially, economists assume that the economy next month will be almost identical to the economy this month, and that any change will be similar to the long-term trend.

One major advantage of this approach, Sims says, is that it brings the hidden assumptions in most models out into the open. Those assumptions, or priors, can be added to the model after identifying the main line of development. "Our feeling is that we have to use models in which controversial assumptions are imposed implicitly, separate from the assumptions needed to stabilize the model and give forecasts," the Minnesota economist adds.

The Sims-Litterman approach, which is called vector autoregression, inevitably arouses skepticism among proponents of the classical model. Klein, for example, says he has decidedly limited expectations from a model that uses "measurement without theory." For very short periods, the vector autoregression model produces good results, Klein says, "but after that it accumulates errors fast."

But almost every econometric model arouses skepticism. One particularly influential line of criticism about forecasts based on economic models comes from the rational-expectations school of economists. These critics say all forecasts are bound to be inaccurate because as soon as they are made public, individuals adjust their economic behavior to make maximum use of the forecast. To give an oversimplified example, if the forecast is for higher meat prices, people will buy less meat, and the reduced demand will keep prices lower than forecast.

Econometric models will always be imperfect, Klein concedes. But weather predictions are also inherently imperfect, yet they are also valuable. It's better to have an imperfect forecast than none at all, Klein adds.

Simulation

Econometric models can ask either "What next?" or "What if?" Another kind of model asks only "What if?" These are called general equilibrium models. They assume that the economy is in a state of equilibrium at any given time and simulate the alterations that would occur as the economy reached a new equilibrium after a major change in some basic factor—taxation or energy prices, for example. These simulation models have a relatively brief history. "In this field, only 15 years ago, we didn't know how to compute economic equilibria," notes John B. Shoven of Stanford University.

The basic question in computing economic equilibrium is, in formal language, whether there exists one set of prices that will clear a market, allowing supply and demand to come into balance. As early as the seventeenth century, Leon Walras of Switzerland made a general statement of the problem.

The first major step in the twentieth century came in the 1930s, when Abraham Wald and John von Neumann outlined mathematical conditions for the

existence of general equilibrium solutions. In the 1950s, Kenneth Arrow and Gerard Debreu—both of whom later won the Nobel Prize—developed a proof more relevant to economists. But while the Arrow-Debreu proof showed that such a set of prices could exist for a general set of models, it gave no guidance on how such market-clearing prices could actually be calculated. "It was like demonstrating that quasars exist but not telling astronomers where they are," Shoven says.

A set of techniques useful for calculating equilibria were developed independently in the 1960s by Herbert Scarf of Yale and Harold Kuhn of Princeton. Those techniques require a large number of calculations, so economic simulations in the late 1960s were limited to what Shoven calls toy models of no more than two or three industries and their customers. Techniques 1,000 to 10,000 times as fast were developed in the early 1970s, and models now can contain much more detail, Shoven says.

He has developed several models. One, designed for studying the effects of changes in tax policy, uses equations describing 19 industries on the production side and 12 different wage-earner income brackets on the consumption side. The U.S. Department of Treasury uses this model in its planning for changes in taxation, Shoven says, and similar models are used by other federal agencies and foreign governments. "The applications have almost exploded," he says.

Miller. Microeconomic model of job turnover.

At Harvard, Dale Jorgenson has developed a model for tax-policy planning and another focused on energy policy. Both place heavy demands on the available computers, which must not only provide access to data in detail—Jorgenson's data base divides labor into 81,000 categories—but also allow the data to be manipulated in different ways. For example, the tax model breaks down commodities into 36 different categories, while the energy model uses only ten categories of commodities.

The models simulating the effects of policy changes are not as advanced as econometric models ("Tax modeling is maybe five or ten years away from that stage," Jorgenson says), but they are perhaps equally confounded by the issue of subjectivity. They require constant assumptions about such basic factors as the effect of a wage increase on the labor supply.

Every model is open to the accusation that an economist has simply fed in the assumptions that will produce a desired result. Yet some backcasting tests, in which efforts were made to reproduce past behavior of the economy, have shown that such models can be highly accurate, Jorgenson says. For example, in a test designed to see why the American economy suddenly took a downturn in 1973, "the model tracked beautifully," says Jorgenson. "It showed that the data have been used accurately in calibrating the model."

Microbehavior

Some subjectivity is inevitable in every area of economic modeling, because the data are always incomplete. For example, microeconomist James J. Heckman of the University of Chicago says that a model of unemployment—designed to show how people look for jobs and why they accept a given job—must work essentially with two pieces of data: how long workers are unemployed, on the average, and what wage they finally accept.

"You can't get readings on all the factors in a process," Heckman explains. "You can only get a reading on the outcome and then build a model to explain those data."

One problem in trying to evaluate such microeconomic models is that they are difficult to subject to the standard scientific test of replicability. The models tend to be very complicated and, Heckman says, "so costly that you can't

Fogel. A million geneologies for 300 years.

ask anyone to duplicate them, so there is a certain amount of skepticism about their value."

For example, a model of job turnover requires that a basic calculation repeat many thousands of times to simulate decisions workers make. Microeconomic models use Bellman equations, developed by Richard Bellman of the Rand Corporation, to simulate sequential decision making over a period of time. Such modeling requires the same kind of calculations needed to simulate melting a block of ice, Heckman says. The rate of melting depends on the temperature around the ice, which changes as the block melts. A good simulation may require 100,000 repetitions of the same solution, each one subtly different from the others, at an enormous cost in computer time.

Despite the complexity of the task, efforts to simulate microeconomic behavior are plentiful. At Carnegie-Mellon University, for example, Robert Miller has developed a microeconomic model of job turnover.

"One of the most striking factors you observe is that the longer a person has been on a job, the less likely he is to quit," Miller notes. There are several explanations for this, he explains. One of them, which has acquired some appeal, is that new employees learn early on whether or not a job matches their expectations. Gaining this knowledge, says Miller, "is most significant in the early stages of a job, so that individual is more likely to leave early on."

On the logical belief that workers find out about some jobs faster than others, Miller worked out the kinds of behavior that might be expected and tested his assumptions against data about job turnover. He then formulated a "fairly specific, highly structured model" using a repetitive system of equations, which generated a probability distribution describing the frequency with which workers can be expected to leave jobs.

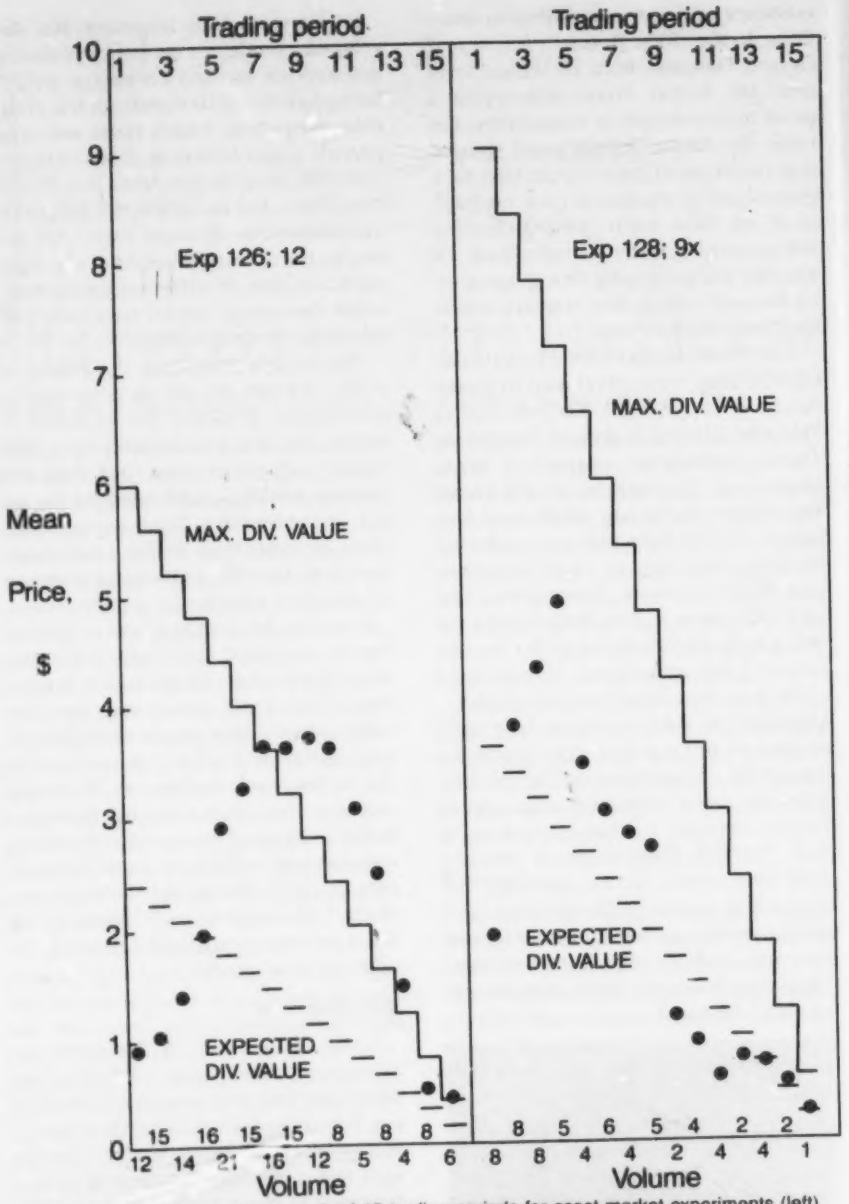
"In this model, you expect young people to turn over more than older people, because they want to learn more about their jobs and themselves," Miller says. "In the past, this was called 'unstable behavior.' What this theoretical model suggests is that it is perfectly reasonable to expect young people to turn over relatively frequently and that older people would be stick-in-the-muds."

Refining the grid

One ultimate goal of such research is to improve macroeconomic modeling by providing more solid data on microeconomic behavior. Perhaps the most ambitious effort in this area involves a group headed by Robert Fogel of the University of Chicago Graduate School of Business. Fogel, working with researchers at a dozen universities and research institutions in the United States and Britain, is building a data base that eventually will include information on 1,000,000 individuals from about 200,000 North American families between 1650 and the present. The data come from genealogies and other sources collected over the past three centuries. The data base will include information on individual wealth, occupation, height, vital statistics, and related characteristics, so as to "understand the microeconomic foundation of macroeconomic behavior," Fogel says.

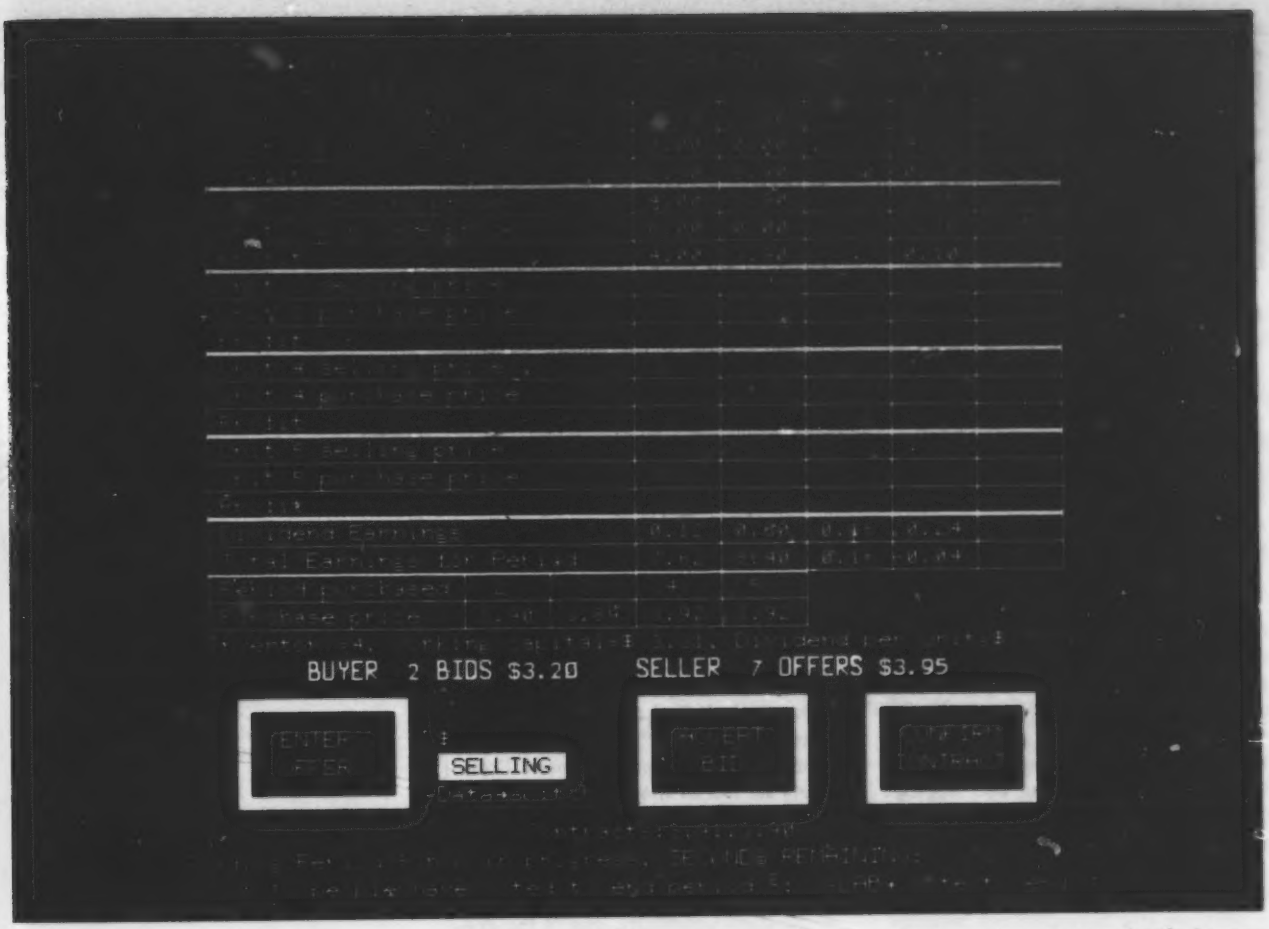
"One purpose is to examine, measure, and explain the decline in mortality that has occurred this century," he says. "The data base is also being used to study the economics of aging, changes in the distribution of wealth, changes in labor force participation rates, migration, and other aspects of economic behavior at the microeconomic level."

Just gathering the information is expected to take ten years, and "it will take a few decades to exploit it," Fogel says. So far, his computers contain data on 72,000 individuals in 10,000 families, linked for up to ten generations. The first results of the project have already



Market trading. Mean price in each of 15 trading periods for asset market experiments (left) and a computer display for a subject in an asset-market experiment (right).

Testing. Vernon Smith (standing) and graduate student simulate market behavior.



been published in papers using height as an indicator of childhood nutrition, correlated to changes in life expectancy.

"There were remarkably high levels of life expectancy at the end of the eighteenth century," Fogel says. "Then there was a decline that lasted for 80 years before the beginning of a new rise, with which we have long been familiar. Our time series on the average stature of native-born white Americans at maturity is strongly correlated with this mortality series.

"Adult height was nearly at the modern level at the time of the American Revolution. There was a decline of about two inches that began with those who were born about 1825. The decline lasted for about 60 years and probably bottomed out in the late 1880s or the early 1890s, when adult heights began to increase again. Our preliminary estimate is that about 40 percent of the long-term decline in mortality was due to changes in nutritional status. Nutritional status refers not merely to diet, but also to claims on the diet that are associated with diseases and other factors. A substantial proportion of the decline in nutritional status was due to increased exposure to diseases that affected physical development during early childhood."

Other studies are aimed at understanding the decline in the U.S. savings rate (from 25 percent in 1890 to 10 percent today) and to quantify the way the wealth of one generation of a family is—or is not—passed on to the next. "It's locked somewhere in the mystery of household behavior," Fogel says.

One interesting new discipline involves a series of laboratory experiments in which individuals recreate some of the real activities of economic life on a small scale. The first such experiments were done in the 1950s and early 1960s, but the field expanded in the 1970s, when Charles R. Plott and Vernon L. Smith began working at the California Institute of Technology. Plott initiated a Caltech program and Smith left to begin experiments at the University of Arizona in 1973.

The experiments are simple in concept, Plott explains. Typically, a small number of volunteers are asked to buy and sell a commodity. Real money is always used to assure the proper incentive. "Essentially, what we're doing is testing the theory of behavior in the market," says Smith.

One finding has been that the form of an auction has a strong influence on prices paid by bidders—or, in more general terms, the sensitivity of markets to market organizations, Plott says. A Dutch auction, in which the price drops by increments until there is a buyer, produces a different result from an English auction, in which buyers bid up the price. Experiments also enable researchers to isolate and analyze the organization of markets more complicated than these.

Bubbles and crashes

"A large series of experiments has demonstrated the fundamental viability of models based on rational-expectations principles," Plott says. Experi-

ments also cast light on some of the most apparently irrational economic behavior. In a series of studies, Smith and Arlington Williams of Indiana University had individuals participate in what is called an oral double auction, the kind used by stock exchanges. "We have been able to produce bubbles and crashes in a laboratory environment," Smith said. "It turned out to be easy."

A bubble, in which the price of an asset is bid up far above any reasonable value before it crashes, is a periodic occurrence in economic history. The most vivid example in this century was the stock market boom that ended in 1929. Smith and Williams have individuals sit at computer terminals and bid with unseen competitors at other terminals for an asset carrying a fixed dividend for a limited time.

According to some economic theories, the price of the asset should fall by the amount of the expected dividend for each time period. The idea of these experiments is to explore the accuracy of competing theories. "In fact, individuals have quite different expectations of what others will do," Smith says. "They bid the price up hoping that another damned fool will take whatever it is off their hands."

Studies have allowed the experimenters to look at appropriate concepts for equilibrium and disequilibrium behavior. They have developed a formula for producing or preventing the bubble/crash cycle. If the experiments include only individuals who have been through the cycle once or twice, the chances of a bubble are small. "But if you have a subset who have not been through a bubble—say five experienced people with three or four inexperienced individuals—you get bubbles and crashes," Smith said.

The results of these experiments have some practical applications. They have been used in regulatory decisions on air freight and inland water barge traffic, and are also being used to explore pricing policies for the planned U.S. space station. The state of Arizona is using them in a program to deregulate electric utilities by setting up an auction that would allow a small number of power distributors to bid for the output of generating stations. Plott and Smith have used their results to suggest that the federal government could handle the problem of airport access by auctioning off landing slots to airlines.

By their nature, laboratory experiments in economics can provide only a sliver of the complexity of the large-scale economy. In theory, economists can use computers to capture some of that complexity by building macroeconomic models on microeconomic foundations. By and large, such modeling remains theoretical because economists do not have access to computers that are powerful enough to handle the computations.

The efforts being made to unite microeconomics and macroeconomics are limited because "there are no bounds to how much computer time you can require," says Christopher Sims. The best hopes lie with a program to set up national supercomputer centers linked in a network. Economists will thus have access to vastly improved computing power. (See "Access to Supercomputers" by Randi Londer, *Mosaic* Volume 16 Number 3.)

"You look at engineers and scientists, and you see a personal computer on every desk, linked to a mainframe next door," Sims says. "Computer power is essentially a free good for them. Economists are not in that situation yet."

While more computer power is on the wish list of economists, it is not necessarily the first item for many of them. Some evidence came when Alan Manne of Stanford University, one of the organizers of a workshop on equilibrium models held in 1984, surveyed the authors of more than a dozen models. He found that only a minority said computing limitations had influenced formulation of their models. And, says Manne, the algorithms now available are generally sufficient for current needs.

What economists do need most of all, says Dale Jorgenson, are "better data and better analyses of existing data."

As it is, Jorgenson says, economic statistics are so incomplete and imperfect that to make them usable in economic models, "the data must be massaged continually." Better data would make economic models easier to use, he declares. So would the availability of more resources for data analysis.

"We'd like to take existing data and do a lot of analysis of it," Jorgenson says. "That's very expensive and takes a lot of scientific talent, which is why it's not done as much as we would like. It would be a great help if we could do that sort of analysis." ●

AVERY PUBLISHING GROUP INC.
89 Baldwin Terrace Wayne, New Jersey 07470

The Mosaic Reader Series

Please send me _____ copies of MOSAIC READER SERIES, at $5.00 each:

[ ] THE CELL [ ] EVOLUTION OF EVOLUTION [ ] ECOLOGY
[ ] DNA [ ] HUMAN EVOLUTION

See inside backcover for details regarding the MOSAIC READER SERIES.

Please send book(s) to:

Name _____

School _____ **Dept** _____

Address _____

City _____ **State** _____ **Zip** _____

Enclosed is my check for $_____. (New York and New Jersey residents add appropriate sales tax.)
Add $1.50 for shipping and handling. Please make check or money order payable to Avery Publishing Group.

Letters, etc.

Honor Out of Chaos

Arthur Fisher, who has written on many subjects for *Mosaic*, won the prestigious 1986 science writing award of the American Institute of Physics for his article "Chaos: The Ultimate Asymmetry," *Mosaic* Volume 16 Number 1. *Mosaic* was also honored by AIP.

Computers to Go with the Flow
Volume 16 Number 4

There was a mismatch of captions with the pictures on page 18 of NCAR's Warren M. Washington and Albert J. Semtner Jr. Here they are correctly identified:

Washington

Semtner

Reader Series

The genius of the biological scientist

As scientists probe the mysteries within our world and universe, **THE MOSAIC READER SERIES** focuses on and highlights the significant current research going on at leading university, industry, and government laboratories. This timely and authoritative series presents to educators, students, and other researchers a concise, uniquely readable overview of the advances being made by preeminent investigators around the world.

AN AUTHORITATIVE SERIES

The first five books available in this fascinating series focus on research in specialized fields of the biological sciences. The volumes in this set are:

THE CELL: Inter- & Intra-Relationships
DNA: The Master Molecule
HUMAN EVOLUTION
EVOLUTION: New Perspectives
ECOLOGY: Impact and Implications

When completed, **THE MOSAIC READER SERIES** will contain more than twenty-four volumes encompassing research in five major areas: Biology, Anthropology, Physics and Astronomy, Chemistry and the Earth Sciences. This comprehensive reference library of current scientific research will examine topics ranging from the topology of DNA and biotechnology to stellar ontogeny and photochemistry.

Each volume is a collection of articles written and edited by some of the nation's most qualified professional science journalists and authenticated by the scientists and engineers engaged in the research being discussed. These articles, originally appearing individually in *Mosaic* magazine, have been carefully evaluated and selected for the breadth and overall lasting significance of the material they cover. Many of the articles contain insets that highlight topics of special interest related to the principal subject. The articles are further enhanced by diagrams, illustrations, and photographs. Each volume has an index and glossary.

To order copies of the titles currently available in **THE MOSAIC READER SERIES**, complete the order card attached at the back of this *Mosaic*.

National Science Foundation
Washington, D C 20550

Official Business
PENALTY FOR PRIVATE USE $300

BULK RATE
POSTAGE & FEES PAID
National Science Foundation
Permit No. G-69

Federal Funds for Mathematics and Computer Sciences Growing Sharply

Federal research support in mathematics and computer sciences has more than quadrupled over the 1976 to 1986 decade, growing from $158 million to $636 million. This increase in funding, averaging nearly 15 percent per year, reflects a growth rate nearly twice that of research funding for all fields of science and engineering combined over that period. In 1986, research support for mathematics and computer sciences is expected to increase 10 percent—the largest relative increase in all the major fields of science and engineering. In contrast, overall Federal obligations for research are expected to decline one percent.

Four agencies receive the most federal research support for mathematics and computer sciences—the Department of Defense (DOD), the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), and the Department of Energy (DOE). Together, they account for 90 percent of all Federal funds to this field.

Over the last decade both basic and applied research in this field have grown rapidly, with basic research funds advancing at an average annual rate of 13.2 percent and applied research, 16.6 percent. During this period there has been an increased emphasis on applied research. In 1976 applied research accounted for approximately 48 percent of the research total, which has grown to 56 percent in 1986. Support for research in mathematics and computer sciences has never been higher, reaching an estimated $282 million in 1986 for basic research funding, and $353 million for applied research funding.

Universities and colleges are heavily relied upon for the performance of research in mathematics and computer sciences. They account for nearly half of all Federal research support to this field—$288 million in 1986. Funding to this sector for mathematics and computer sciences has grown faster than has the total Federal support to this field. Funds have increased an average annual rate of 17.5 percent over the 1976-86 decade. DOD and NSF account for virtually all of the research support to academia in mathematics and computer sciences.

About two-thirds of these funds to academia, a total of $189 million in 1986, are for basic research. Growth in academic basic research has been considerable since 1976, with funds increasing by 15.1 percent per year. Applied research funds to academia have grown much faster, at an average annual rate of 24.3 percent. Academic applied research funding in this field amounts to $99 million in 1986, with DOD providing almost all of these funds.

Since 1976 the fraction of DOD and NSF funds allocated to basic research, as distinct from applied research, in mathematics and computer sciences has remained relatively stable. DOD has allocated approximately 70 percent of its 1986 funds to applied research and 30 percent to basic, compared with a ratio of 60 percent to 40 percent in 1976. NSF places virtually all of its support in basic research. Some shifts have occurred over the decade in NASA and DOE, which have increasingly emphasized applied research in recent years. NASA has increased its ratio of support for applied research from 25 percent to 60 percent, while DOE has increased from solely basic research support to nearly 40 percent applied research funding.

Distribution of federal research obligations for mathematics and computer sciences: FY 1976 and FY 1986 (est.)